

2. KIADÁS FRISSÍTVE
ÉS BŐVÍTVE

DATA DRIVEN CONSTRUCTION

AZ ADATKORSZAK ELIGAZODÁSA AZ
ÉPÍTŐIPARBAN

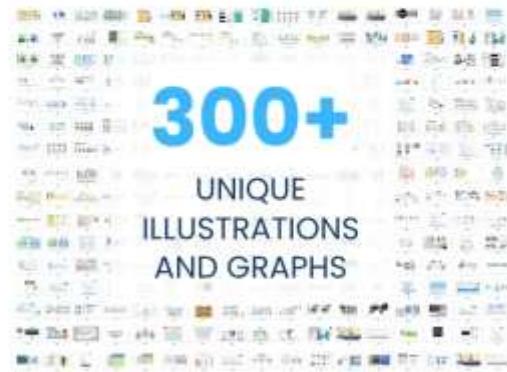
AI ÉS LLM HASZNÁLATI
ESETEKKEL

Artem Boiko



50+

PRACTICAL
PROBLEMS
SOLVED



300+

UNIQUE
ILLUSTRATIONS
AND GRAPHS

DATA-DRIVEN CONSTRUCTION

Navigálás az adatkorszakban
az építőiparban

Második, átdolgozott és kiegészített kiadás

ARTEM BOIKO

“

"Boiko az informatika James Carville-je - utóbbi sokat idézett "It's the economy, stupid" (A gazdaságról van szó, te hülye) című mondatában csak egy szót kell kicserélni erre a híres könyvre. "It's the data, stupid". (Az adatuniverzumban való eligazodáshoz pedig ma is érvényes az ókori rómaiak görög eredetű mondása: "Navigare necesse est". A szerző biztos kézzel és rendíthetetlen iránytűvel navigálja olvasóit az adatóceán minden mélységében és sekély-ségében, nem beszélve az átfogó történeti megközelítésről és nem utolsósorban a rendkívül eredeti grafikákról és a nem csak második pillantásra feltűnő jó humorról. Boiko könyvének nemzetközi visszhangja az eufórikus helyesléstől a meglehetősen epés szkepticismusig terjed, ami jót tett a könyv második német kiadásának. Boiko eredeti és ésogmatikus adatgondolkodó. Izgalmas meglátásokkal és mindig bátor, sőt provokatív, továbbgondolásra ösztönző tézisekkel ajándékozza meg az olvasót. Kiváló gyógyszer a látens konszenzualizmus német betegségére. A fenti latin közmanodásnak egyébként van egy kiegészítése: "vivere non est necesse". Ez nem vonatkozik Boiko megközelítésére az adatok világához - az adatok élnek, és az életük szükséges, hogy ne mondjam, döntő fontosságú."

- Dr. Burkhard Talebitari, szabadúszó szerkesztő - többek között az Ernst & Sohn által 2013 óta évente kiadott BIM című folyóirat számára.

"Artem Boiko könyve mérföldkő a digitalizáció demokratizálása szempontjából az építőiparban - és valódi változást jelent a kis- és középvállalkozások (kkv-k) számára. Különösen úttörő: a modern, nyílt forráskódú, alacsony kód nélküli eszközök használatával a vállalatok már most is hatékonyan integrálhatják az adatokat üzleti folyamataikba és nyereségesen elemezhetik azokat - mélyreható programozási ismeretek nélkül. Ez feleslegessé teszi a nehézsékes kereskedelmi szoftvercsomagok költséges használatát. Ez a könyv felhívás a cselekvésre! Értékes útmutató mindenek számára, akik nemcsak megérteni akarják az építőipari digitális átalakulását, hanem aktívan alakítani is akarják azt - pragmatikusan, hatékonyan és előremutatóan. Itt az ideje, hogy együtt dolgozzunk ezen ismeretek megesztésén és az építőipari termelékenységének fenntartható növelésén."

- **Dr. Michael Max Buehler**, a konstanzi HTWG Konstanz építőipari menedzsment professzora, a GemeinWerk Ventures társtulajdonosa és a DevvStream független igazgatója.

"DataDrivenConstruction könyv az egyik első lépést túl a határain a szokásos világ építői, a komplex tervezési és irányítási rendszerek, amikor, úgy tűnik, a komplexitás és telítettség az adatok nem is ad esélyt a radikális egyszerűsítés és a fokozott átláthatóság a munka építési adatokkal. Artem könyvében egyszerű nyelven mutatja be, hogy az adatokkal való munka modern technológiái minden lehetőséget nyitnak meg előttünk, és szó szerint konkrét lépésekkel ad, amelyeket azonnal alkalmazhatunk a munkánk során. Mindenkit arra kérlek, aki meg akarja érteni, hogy az automatizálási rendszerek merre fognak haladni az építőiparban, hogy figyelmesen tanulmányozza ezt a könyvet, hogy felismerje: az adatforradalom az építőiparban már kopogtat az ajtónkon. Ez most még csak a geekeket érdekel, de néhány éven belül, akárcsak a BIM, az ilyen megközelítések és szoftverek mindenütt jelen lesznek!"

- **Ihor Rogachew**, az IMT Kompetencia Központ vezetője, BIM és digitális átalakulás az RGD-nél, valamint az InfraBIM.Pro alapítója.

"Nagyon ajánlom a DataDrivenConstruction könyvet, amely - ahogy a cím is mondja - az AECO adatvezérelt információkezelési megközelítésével foglalkozik. Jelenleg arra használom, hogy segítsen számos megbeszélést kezdeményezni különböző csoportokkal. Nagyon hozzáférhető referenciának találtam. Az AECO eszközeinek történeti háttérének alapos áttekintése, az adatok és számos kulcsfontosságú technológia bemutatása mellett a könyv számos nagyon hasznos ábrát tartalmaz, amelyek az adatforrások és a végfelhasználói artefaktumok körét vázolják fel, minta-munkafolyamatokkal. Úgy tűnik nekem, hogy az ilyen típusú diagramokra van szükségünk többet az információs stratégiák kidolgozása és nyomon követése során, és hozzájárulnak a BEP-ekhez - meghatározzák a teljes vállalati adatmodellt, amelyre a PIM és az AIM határai ráteríthetők."

- **Paul Ransley**, az Acmena vezető tanácsadója és a Transport for London rendszerintegrációs mérnöke.

"Ha "az adat az új olaj", akkor meg kell tanulnunk meghatározni, megtalálni, bányászni, finomítani, hogy értékessé tegyük. Nagyon informatívnak és tanulságosnak találtam a DataDrivenConstruction című könyvet. A könyv hasznos történelmi háttérrel nyújt, és közérthető nyelven magyarázza el az adatokkal való munkát. Azok számára, akiket érdekel a digitális átalakulás, jó megértést ad az adatokról - hogyan működnek, hogyan strukturálódnak és hogyan használhatók fel."

- **Ralph Montague**, az ArcDox igazgatója, a BIM koordinátorok csúcstartalozójának igazgatója és az írországi Nemzeti Szabványügyi Hatóság BIM Nemzeti Tükörbizottságának elnöke.

"Amint azt a könyvben is hangsúlyozták, az információ az építőipari ágazat számára alapvető fontosságú, és a hozzáérhető formátumban való rendelkezésre állás nagymértékben megkönyíti a pontos döntéshozatalt és felgyorsítja a projektek ütemezését. A könyv semleges és hatékony megközelítést kínál a forrás eléréséhez és kihasználásához a döntéshozatalban. A könyvben bemutatott módszertan olyan kortárs megközelítést használ, amely a mesterséges intelligencia által vezérelt programozást és a hozzáférhető nyílt forráskódú eszközöket ötvözi. Az AI erejének kihasználásával és a nyílt forráskódú szoftverek felhasználásával a módszertan célja az automatizálás fokozása, a folyamatok optimalizálása, valamint a szakterületen belüli hozzáférhetőség és együttműködés elősegítése. A könyv nyelvezete világos és könnyen követhető."

- **Dr. Salih Ofluoğlu**, az Antalya Bilim Egyetem Képzőművészeti és Építészeti Karának dékánja és az Eurázsiai BIM Fórum szervezője.

"Csak annyit tudok mondani, hogy WOW! Az a mód, ahogyan beépítette a történelmet, az LLM, a grafikákat és a pontok megértésének általános könnyedségét, valóban figyelemre méltó. A könyv áramlása elköpesztő. Annyi ragyogó aspektusa van ennek a könyvenek; ez valóban egy játékváltó. Ez egy nagyszerű információforrás, és elismerésemet fejezem ki az erőfeszítésért és a szenvédélyért, amit beleölt. Gratulálok, hogy ilyen figyelemre méltó művet alkottál. Folytathatnám még, de elég, ha csak annyit mondok, hogy hihetetlenül le vagyok nyúgözve!"

- **Natasha Prinsloo**, az energylab_ digitális gyakorlatának vezetője.

"Az építőiparban dolgozók számára, a kezdőtől a tapasztalt profikig, ez a könyv megváltoztatja a játékot! Nem egy tipikus poros olvasmány - tele van meglátásokkal, stratégiákkal és egy csipetnyi humorral, hogy lekössé az embert. Az ósi adatrögzítési módszerektől a legmodernebb digitális technológiákig, az építőiparban történő adatfelhasználás fejlődését mutatja be. Olyan, mintha egy időgéppel utazna az építési adatok evolúcióján keresztül. Akár építész, mérnök, projektmenedzser vagy adatelemző, ez az átfogó útmutató forradalmasítani fogja a projektek megközelítését. Készüljön fel a folyamatok optimalizálására, a döntéshozatal fokozására és a projektek olyan irányítására, mint még soha!"

- **Pierpaolo Vergati**, a Római Sapienza Egyetem oktatója és a Fintecna vezető építési projektmenedzsere.

"Egy lélegzetvétel alatt, kevesebb mint 6 óra alatt olvastam el a könyvet. A könyv gyártási minősége kiváló, sűrű fényes papír, színvilág, kellemes betütípus. A nagyszámú gyakorlati példa az építőiparra jellemző LLM munkamódsszerrel való munkavégzésre hónapok, ha nem évek önképzésétől kiméli meg az olvasót. A munkapéldák nagyon változatosak, az egyszerűtől a bonyolultig terjednek, anélkül, hogy bonyolult és drága szoftverek beszerzésére lenne szükség. A könyv lehetővé teszi, hogy az építőiparban működő bármely vállalkozás tulajdonosai új szemmel tekintsenek üzleti stratégiájukra, digitalizációjukra és fejlesztési kilátásaiakra. A kisebb cégek számára pedig azt, hogy megfizethető és ingyenes eszközökkel növeljék a hatékonyságot."

- **Mikhail Kosarev**, a TIM-ASG előadója és tanácsadója az építőipar digitális átalakulásáról.

A "DATA DRIVEN CONSTRUCTION" könyv egy olyan könyv, amely megváltoztatja a játékot mindenek számára, akik kíváncsiak arra, hogy az építőipar merre tart az adatok korában. Artem nem csak a felszínt kapargatja; mélyen belemerül az építőipar jelenlegi fejleményeibe, kihívásaiba és igéretes lehetőségeibe. A könyv különlegessége a hozzáérhetőségében rejlik - Artem az összetett gondolatokat olyan közérthető analógiákkal magyarázza el, amelyek segítségével a tartalom könnyen megérthetővé válik. A könyvet hihetetlenül informatívanak, ugyanakkor magával ragadónak találtam. Összefoglalva, Artem egy olyan értékes forrást alkotott, amely nemcsak tájékoztat, hanem inspirál is. Akár tapasztalt szakember, akár az építőipar újdonsült szereplője, ez a könyv szélesít a látókörét, és elmélyíti a megértését azzal kapcsolatban, hogy merre tart az iparág. Nagyon ajánlott!"

- **Moayad Saleh**, építész és a TMM GROUP Gesamtplanungs GmbH BIM implementációs vezetője.

"Azt kell mondanom, hogy az Adatvezérelt építés méltó arra, hogy tankönyvként tanítsák az egyetemeken, és olyan könyv, amely értékes hozzájárulást fog nyújtani a fejlődő BIM területéhez. Az Adatvezérelt építés könyv tartalmaz egy technikai szójegyzéket, amely nagyon jól magyarázza a fogalmakat. A rendkívül nehezen magyarázható témákat nagyon szép vizuális nyelvezettel teszik egyszerűvé és érthetővé. Úgy gondolom, hogy amit a vizuális anyagban meg akarnak magyarázni, azt az olvasónak, ha röviden is, de ki kell fejteni. Egyes vizuálisok érthetősége, vagyis a vizuális olvasása külön információt igényel. Szeretném azt is elmondani, hogy szívesen mutatom be Artem Boiko értékes munkáit az egyetemi előadásaimon és szemináriumaimon."

- **Dr. Ediz Yazıcıoglu**, az ArchCube tulajdonosa, valamint az Isztambuli Műszaki Egyetem és a Medipol Egyetem építészeti tanszékének építési projektmenedzsment előadója.

"Az adatvezérelt építés szemléletesen közvetíti az építési adatokkal való információalapú munka alapjait. Az információáramlást és az alapvető gazdasági fogalmakat tárgyaló könyv, és ezzel kiemelkedik a többi BIM könyv közül, mert nemcsak egy szoftvergyártó szemszögét képviseli, hanem alapvető fogalmakat is igyekszik közvetíteni. Egy könyv, amelyet érdemes elolvasni és megnézni."

- **Jakob Hirn**, a Build Informed GmbH vezérigazgatója és társalapítója, valamint a "On Top With BIM" innovációs fórum kezdeményezője.

"Az adat az új olaj", ahogy mondani szokták, ezért a kutatóknak vagy bányászoknak megfelelő eszközökkel és gondolkodásmóddal kell rendelkezniük ahoz, hogy értéket nyerjenek ebből a 21. századi erőforrásból. Az építőipar már túl régóta a "3D-információ" alapuló folyamatok csúszós lejtőjén jár, ahol a projekt megvalósítása valaki más által sütött információkon alapul (pl. már felrajzolták a kör- vagy oszlopdiagramot), miközben a mögöttes "adatok" (pl. a nyers táblázat) sokkal többet tudnak nyújtanı, különösen azért, mert a több adat fúziója és a mesterséges intelligencia korlátlan lehetőségeket rejt magában. Ha Ön építőipari kivitelezést végez (vagy tanít/kutat), ez a könyv a legjobb - és eddig egyetlen - forrása az adatvezérelt világban való eligazodáshoz, amelyre most találtunk rá."

- **Dr. Zulfikar Adamu**, az építőipari stratégiai informatika docense az LSBU-n (Egyesült Királyság).

Az Artem Boiko által írt "Data-Driven Construction" lenyűgöző munka, amely szilárd alapokat kínál az építőipar számára a folymatosan növekvő technológiák és információs lehetőségek idején. Boikónak sikérül az összetett témaákat érthetően bemutatnia, miközben látomásos ötleteket is bemutat. A könyv egy jól átgondolt kompendium, amely nemcsak a jelenlegi fejleményekre világít rá, hanem a jövőbeli újításokra is kitékítést nyújt. Mindenkinek erősen ajánlott, aki szeretne megismerni az adatvezérelt építőipari tervezéssel és kivitelezéssel."

- **Markus Eiberger**, a Stuttgarti Alkalmazott Tudományok Egyetemének oktatója, a Konstruktionsgruppe Bauen vezető projektmenedzsere és kirendeltségvezető-helyettese, a BIM Cluster Baden-Württemberg Egyesület elnökségi tagja.



Második kiadás, 2025. március.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-04-1



Artem Boiko Szerzői jog

boikoartem@gmail.com
info@datadrivenconstruction.io

E könyv egyetlen része sem sokszorosítható vagy továbbítható semmilyen formában vagy bármilyen elektronikus vagy mechanikus eszközzel, beleértve a fénymásolást, a rögzítést vagy bármilyen információtároló és -kereső rendszert, a szerző írásos engedélye nélkül - kivéve a nem kereskedelmi célú terjesztést változatlan formában. A könyvet ingyenesen terjesztjük, és szabadon megoszthatjuk más felhasználókkal személyes, oktatási vagy kutatási célokra, feltéve, hogy a szerzőség és az eredetire való hivatkozás megmarad. A szerző fenntartja a szöveggel kapcsolatos minden nem tulajdonosi jogot, és nem vállal sem kifejezetten, sem hallgatólagos garanciát. A könyvben említett vállalatok, termékek és nevek fiktívek vagy példaként használhatók. A szerző nem vállal felelősséget a közölt információk felhasználásának következményeiért. A könyvben szereplő információk "úgy, ahogy van", a teljességre és a helytállóságra vonatkozó garancia nélkül kerülnek átadásra. A szerző nem vállal felelősséget a könyvben szereplő információk, kódok vagy programok használatából eredő véletlen vagy következményes károkért. A könyvben bemutatott kódminták kizárálag oktatási célokat szolgálnak. Az olvasó saját felelősségeire használja őket. A szerző azt ajánlja, hogy minden szoftvermegoldást teszteljenek le, mielőtt azt termelési környezetben használnák. A szövegben említett valamennyi védjegy és terméknév a megfelelő vállalatok védjegye, bejegyzett védjegye vagy szolgáltatási védjegye, és a megfelelő tulajdonosok tulajdonát képezi. Ezeknek a neveknek a könyvben való használata nem jelenti azt, hogy a tulajdonosokkal bármilyen kapcsolatban állnak vagy azokat támogatnák. Harmadik fél termékeinek vagy szolgáltatásainak említése nem minősül ajánlásnak, és nem is jelent jóváhagyást. A példákban használt cég- és terméknevek a tulajdonosok védjegyei lehetnek. A harmadik felek weboldalaira mutató linkek csak kényelmi céllal szerepelnek, és nem jelentik azt, hogy a szerző támogatja az ott található információkat. minden idézett statisztika, idézet és tanulmány a cikk megírásának időpontjában volt aktuális. Az adatok idővel változhatnak.

Ez a könyv a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licenc (CC BY-NC-NC-ND 4.0) alatt kerül terjesztésre. Nem kereskedelmi célokra másolhatja és terjesztheti, feltéve, hogy megtartja a szerzőséget, és nem változtat rajta.



CC BY-NC-ND 4.0

© 2024 Artyom Boiko. Első kiadás.
© 2025 Artyom Boiko. Második, átdolgozott és kiegészített kiadás.
Minden jog fenntarta.

ELŐSZÓ A MÁSODIK KIADÁSHOZ

Ez a könyv a szakmai közösséggel folytatott élénk párbeszéd eredménye. Az építőipari adatkezelésről szóló számos szakmai vitán alapul, amelyek különböző szakmai platformokon és közösségi médiafelületeken zajlottak. Ezek a viták cikkek, kiadványok és vizuális anyagok alapjául szolgáltak, amelyek széles körű visszhangot váltottak ki a szakmai közösségen. A szerző tartalma évente több millió megtékinést vonz a különböző platformokon és nyelveken, és összehozza az építőipar digitalizációjával foglalkozó szakembereket.

Az első kiadás megjelenését követő egy éven belül a könyvet több mint 50 ország szakértői rendelték meg, Brazíliától és Perutól Mauritiusig és Japánig. A könyv második kiadását, amelyet most Ön a kezében tart, a szakértői visszajelzések, az első kiadással kapcsolatos kritikák és a szakmai körökben folytatott viták alapján átdolgozták és kibővítették. A visszajelzéseknek köszönhetően a második kiadás jelentősen bővült: új fejezetek kerültek be a CAD (BIM) technológiákról és a hatékony ETL -folyamatok létrehozásáról. A gyakorlati példák és esettanulmányok száma is jelentősen megnőtt. Különösen értékes az építőipar, a tanácsadó cégek és a nagy IT-vállalatok vezetőinek visszajelzése, akik a könyv első változatának megjelenése előtt és után is a szerzőhöz fordultak a digitalizálással és az interoperabilitással kapcsolatos kérdésekkel. Közülük sokan már alkalmazták a könyvben leírt megközelítéseket, vagy a közeljövőben tervezik ezt megtenni

Ön egy olyan könyvet tart a kezében, amely vita és aktív véleménycsere révén jött létre. A fejlődés a párbeszédben, a nézetek ütközötésében és az új megközelítésekre való nyitottságban születik. Köszönjük, hogy részt vesz ebben a párbeszédben. Az Ön konstruktív kritikája a jövőbeli fejlesztések alapja. Ha hibákat talál a szövegben, vagy ötleteket és javaslatokat szeretne megosztani velünk, minden visszajelzést szívesen fogadunk. Az elérhetőségek a könyv végén találhatók

MIÉRT INGYENES A KÖNYV?

Ez a könyv nyílt oktatási forrásként készült, amelynek célja az építőiparban az adatkezelés modern megközelítéseinek terjesztése. A könyv első változata szolgált alapul a szakmai közösség észrevételeinek és javaslatainak összegyűjtéséhez, ami lehetővé tette az anyag szerkezetének és tartalmának javítását. minden megjegyzést, javaslatot és ötletet gondosan elemezünk és beépítettük ebbe az átdolgozott változatba. A könyv célja, hogy segítsen az építőipari szakembereknek megérteni az adatokkal való munka fontosságát: szisztematikusan, tudatosan és az információk hosszú távú értékét szem előtt tartva. A szerző az építőipar digitalizásával kapcsolatos több mint 10 éves munkája során gyűjtött össze példákat, illusztrációkat és gyakorlati észrevételeket. Az anyag nagy része valós projektekből, mérnökökkel és kivitelezőkkel folytatott beszélgetésekből, nemzetközi kezdeményezésekben és képzési szemináriumokon való részvételből született. A könyv kísérlet arra, hogy a felhalmozott tapasztalatokat rendszerezze és közérthető módon megossza. Ha szeretné támogatni a könyvben megfogalmazott gondolatok további terjesztését, és kényelmes formátumban szeretne olvasni, példákkal és képanyagokkal dolgozni - megvásárolhatja [a nyomtatott változatot](#).

HASZNÁLATI JOGOK

A könyv minden anyaga, illusztrációja és töredéke bármilyen formában és bármilyen adathordozón sokszorosítható, idézhető vagy felhasználható, feltéve, hogy feltüntetik a forrást: Artem Boiko szerzősége és a könyv címe "Adatvezérelt építés". Köszönjük a munka tiszteletét és a tudás terjesztését.

Őszinte hálával ajánlom ezt a könyvet a családomnak, akik már korán belém nevelték az építőipar iránti mély szeretetet, a bányászvárosomnak, ahol az ellenálló képességet tanultam, és földmérő feleségemnek, akinek rendíthatetlen támogatása állandó inspirációm volt.

KINEK SZÓL EZ A KÖNYV

A közérhető nyelven megírt könyv az építőiparban dolgozó olvasók széles körét célozza meg - a diákoktól és a kezdőktől kezdve, akik a modern építési folyamatok alapjait szeretnék megismerni, egészen a szakemberekig, akiknek az építőipari adatkezelés naprakész módszertanára van szükségük. Legyen Ön építész, mérnök, művezető, építésvezető vagy adatelemző, ez a számos egyedi illusztrációval és grafikával ellátott átfogó útmutató értekes betekintést nyújt abba, hogyan lehet az adatokat az üzleti életben a folyamatok optimalizálására és automatizálására, a döntéshozatal javítására és az építési projektek különböző szinteken történő, modern eszközökkel történő irányítására használni.

A könyv átfogó útmutató, amely egyesíti az elméleti alapokat és a gyakorlati ajánlásokat az adatkezelési technikák építési folyamatokba történő integrálásához. A könyv az információk stratégiai felhasználására összpontosít a műveletek optimalizálása, a folyamatok automatizálása, a döntéshozatal javítása és a projektek hatékony irányítása érdekében a modern digitális eszközök segítségével.

Ez a könyv az építőiparban az információval való munka elméleti és gyakorlati szempontjait tárgyalja. Részletes példákon keresztül vizsgálja a feladatok paraméterezésének, a követelmények összegyűjtésének, a strukturáltalan és több formátumú adatok feldolgozásának és az építőipari vállalatok számára hatékony megoldásokká alakításának módszertanát.

Az olvasó fokozatosan halad a követelmények kialakításától és az alapvető adatmodellek kidolgozásától a heterogén információforrások integrációjának összetettebb folyamataiig, az ETL-folyamatok létrehozásáig, az információs csővezetékek és a gépi tanulási modellek felépítéséig. A szekvenciális megközelítés lehetővé teszi az építőiparban az üzleti folyamatok és a döntéstámogató rendszerek szervezésének és automatizálásának mechanizmusainak világos bemutatását. A könyv minden részét egy gyakorlati fejezet zárja, amely lépésről-lépére követhető utasításokat tartalmaz, amelyek lehetővé teszik a megszerzett tudás azonnali alkalmazását valós projektekben.

A KÖNYV EGYES RÉSZEINEK ÖSSZEFoglalása

Ez a könyv az értékláncban történő adatátalakítás koncepciója köré épül: az adatgyűjtéstől és minőségbiztosítástól az analitikus feldolgozásig, valamint az értékes gyakorlati megoldások kinyeréséig, modern eszközök és módszertanok használatával.

1. rész: Digitális fejlődés az építőiparban - az adatkezelés történelmi átalakulását követi nyomon az agyagtábláktól a modern digitális rendszerekig, elemezve a moduláris rendszerek megjelenését és az információs digitalizáció növekvő jelentőségét az ipari forradalmak összefüggésében.

2. rész: Információs kihívások az építőipar számára - az adatok széttöredezetségének, az "információs silóknak", a HiPPO megközelítésnek a döntéshozatalra gyakorolt hatásának és a védett formátumok korlátainak a problémáit vizsgálja, javasolva a mesterséges intelligencia és az LLM ökoszisztémák felé való elmozdulás megfontolását.

3. rész: Az építőipari adatok rendszerezése - az építőipari adatok tipológiáját alkotja, leírja az adatok rendszerezésének módszereit, a vállalati rendszerekkel való integráját, és tárgyalja az információs folyamatok szabványosítására szolgáló kompetencia központok létrehozását.

4. rész: Adatminőség-ellenőrzés - feltárja azokat a módszereket, amelyekkel az eltérő információkból minőségi, strukturált adatokat lehet készíteni, beleértve az adatok különböző forrásokból történő kinyerését, validálását és modellezését az LLM segítségével.

5. rész: Költség- és időszámítások - a költség- és tervszámítások digitalizálásával, a CAD (BIM) modellekből történő térfogatnyerés automatizálásával, a 4D-8D modellezési technológiákkal és az építési projektek ESG-számításával foglalkozik.

6. rész: CAD és BIM - kritikusan elemzi a tervezési technológiák fejlődését, a rendszerek interoperabilitásának kérdéseit, a nyílt adatformátumok irányába mutató tendenciákat és a mesterséges intelligencia tervezésre való alkalmazásának kilátásait.

7. rész: Adatelemzés és automatizálás - az információ vizualizációjának alapelveivel, a kulcsfontosságú teljesítménymutatókkal, az ETL-folyamatokkal, a munkafolyamat-rendszerező eszközökkel és a nyelvi modellek alkalmazásával foglalkozik a rutinfeladatok automatizálása érdekében.

8. rész: Adattárolás és -kezelés - az adattárolási formátumok, az adattárház és az adathalmaz fogalmak, az adatkezelési elvek és az új megközelítések, köztük a vektoradatbázisok, valamint a DataOps és VectorOps módszertanok feltárása.

9. rész: Big Data és gépi tanulás - a történelmi adatokon alapuló objektív elemzésre való áttérésre, a dolgok internetére az építkezések, valamint a gépi tanulási algoritmusok alkalmazására a projektköltségek és határidők előrejelzésére összpontosít.

A 10. rész: Az építőipar a digitális adatok korában - az építőipar jövőjét mutatja be, elemezve az ok-okozati elemzésről az összefüggésekkel való munkára való áttérést, az építőipar "Uberizációjának" koncepcióját és a digitális átalakulás stratégiáit.

What is meant by **data-driven construction** ?



BEVEZETÉS

Meddig maradhat versenyképes az Ön cége egy olyan világban, ahol a technológia rohamosan fejlődik, és az üzlet minden aspektusát, az időzítéstől és a költségszámítástól a kockázatelemzésig, automatizálja a gépi tanulási modellek?

Az építőipar, amely az emberiséggel egyidős, forradalmi változások küszöbén áll, amelyek teljesen megváltoztatják a hagyományos építőiparról alkotott elképzeléseinket. A gazdaság más ágazataiban a digitalizáció már most is nemcsak a szabályokat változtatja meg, hanem kíméletlenül kiszorítja a piacról azokat a vállalatokat, amelyek nem tudtak alkalmazkodni az új adatfeldolgozási környezethez, és nem képesek javítani a döntéshozatal sebességét (1. ábra).



1. ábra Az építőiparban a döntéshozatal sebessége gyakrabban függ az emberi tényezőtől, mint más iparágakban.

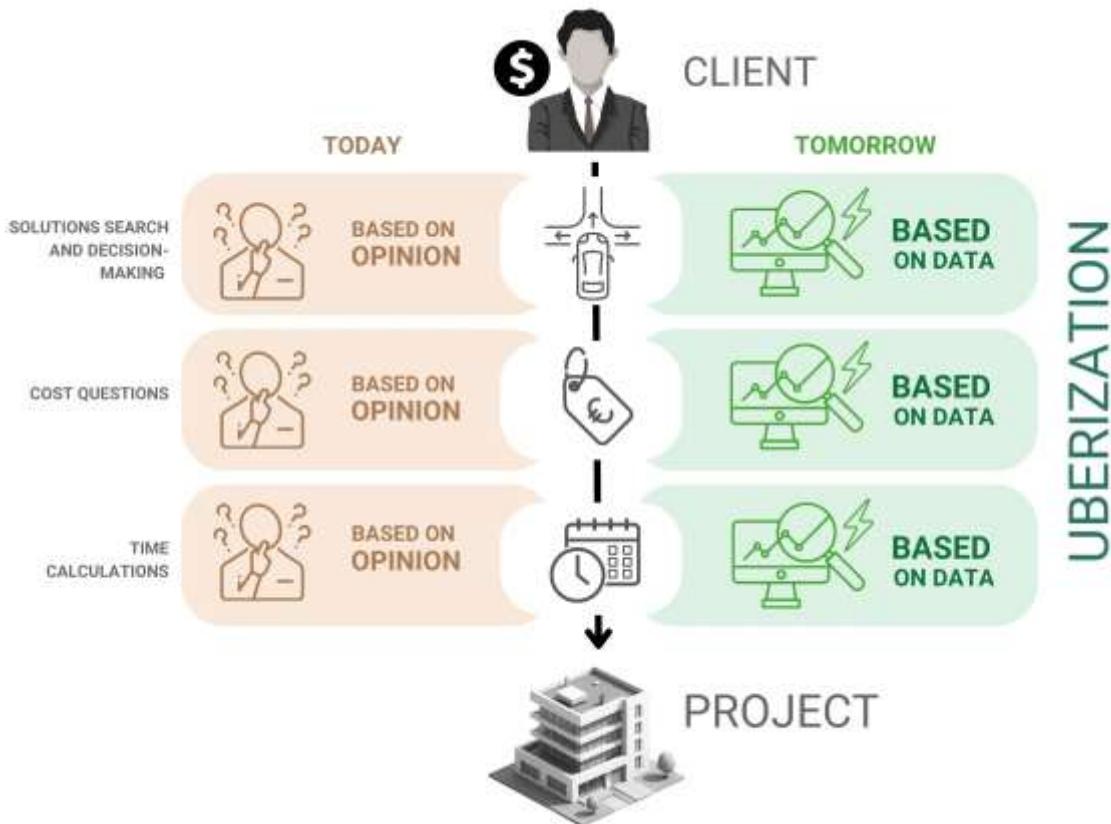
A bankszektor, a kiskereskedeleml, a logisztika és az agrárgazdaság gyorsan halad a teljes digitalizáció felé, ahol a pontatlanságoknak és a szubjektív véleményeknek már nincs helye. A modern algoritmusok hatalmas mennyiséggű adatot képesek elemezni, és pontos előrejelzéseket nyújtani az ügyfeleknek - legyen szó a hitel visszafizetésének valószínűségéről, az optimális szállítási útvonalakról vagy a kockázat előrejelzséről.

Az építőipar az egyik utolsó iparág, amelyik elkerülhetetlenül átáll a magasan fizetett szakemberek véleményén alapuló megoldásokról az adatvezérelt megoldásokra. Ezt az átmenetet nemcsak az új technológiai lehetőségek, hanem az átláthatóság, a pontosság és a gyorsaság iránti megnövekedett piaci és ügyféligeny is vezérli.

Robotizáció, folyamatautomatizálás, nyílt adatok és az ezeken alapuló előrejelzések - mindenek már nem csak lehetőségek, hanem elkerülhetetlenek. Az építőiparban működő legtöbb vállalat, amelyek nemrégiben még a megrendelőnek feleltek a projektek mennyiségének, költségének, idejének és

minőségellenőrzésének kiszámításáért, ma már azt kockáztatják, hogy pusztán a megrendelések végrehajtójával válnak, és nem hoznak kulcsfontosságú döntéseket (2. ábra).

A számítási teljesítmény, a gépi tanulási algoritmusok és az adathozzáférés demokratizálódásával ma már lehetséges a különböző forrásokból származó adatok automatikus kombinálása, ami mélyebb folyamatelemzést, kockázat-előrejelzést és költségoptimalizálást tesz lehetővé az építési projekt tárgyalási szakaszaiban. Ezek a technológiák képesek radikálisan javítani a hatékonyságot és csökkenteni a költségeket az egész ágazatban.



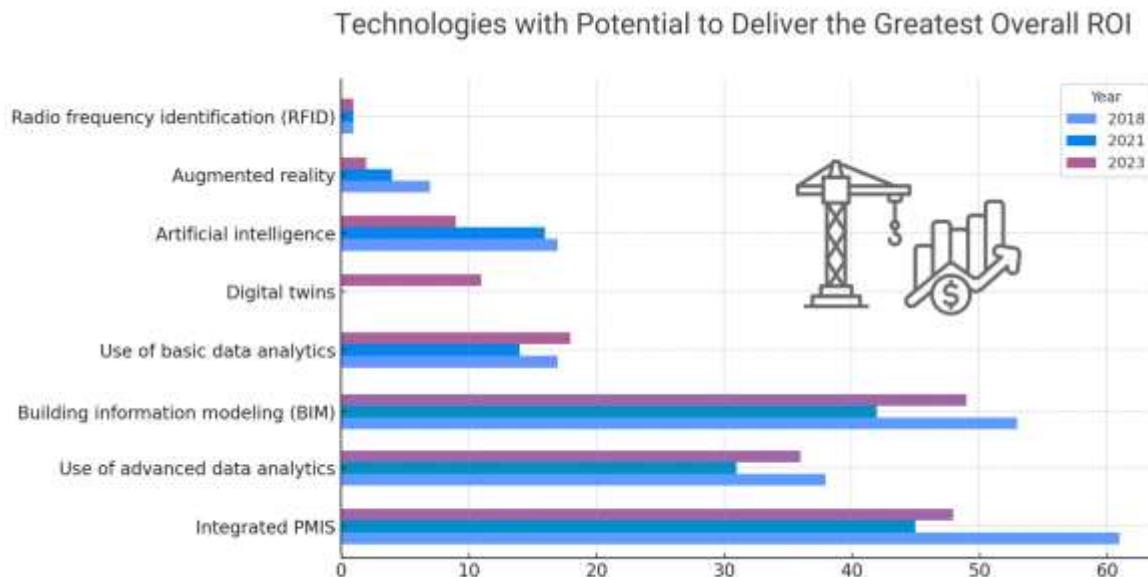
2. ábra Az ügyfelet nem érdeklik a túlzott emberi tényezők a projekt megvalósítása során.

Az új eszközök és koncepciók minden előnye ellenére az építőipar messze elmarad a gazdaság más ágazatai mögött az új technológiák átvételében.

Az IT Metrics Key Data 2017 jelentés szerint az építőipar az utolsó helyen áll az informatikai kiadások terén 19 másik iparág közül [1].

Az adatmennyiség és a folyamatok összetettségének gyors növekedése egyre nagyobb fejfájást okoz a vállalatvezetésnek, és az új technológiák használatának fő problémája az, hogy az adatok - bőségük ellenére - töredzettek, strukturálatlanok és gyakran nem kompatibilisek a különböző rendszerek és szoftvertermékek között. Ezért az építőipari ágazatban számos vállalatot ma már elsősorban az adatminőségi kérdések foglalkoztatnak, amelyek csak hatékony, automatizált irányítási és elemzési rendszerek bevezetésével oldhatók meg.

A KPMG® által az építésügyi vezetők körében 2023-ban végzett felmérés [2] szerint a projektirányítási információs rendszerek (PMIS), a fejlett és alapvető adatelemzés és az épületinformációs modellezés (BIM) rendelkezik a legnagyobb potenciállal a projektek megtérülésének javítására (3. ábra).



3. ábra Az építőipari vállalatok vezetői körében végzett felmérés: mely technológiák biztosítják a beruházás legnagyobb megtérülését (ROI) a beruházási projekteknél? (a [2] anyag alapján).

Az adatok üzleti folyamatokba történő integrálásával kapcsolatos kihívások megoldása a magas minőségű információk biztosítása, a megfelelő adatformátumok használata, valamint az adatok létrehozására, tárolására, elemzésére és feldolgozására szolgáló hatékony módszerek alkalmazása.

Az adatok értékének felismerése arra kényszeríti a különböző iparágakat, hogy eltávolodjanak a silózott alkalmazásoktól és a bonyolult bürokratikus irányítási struktúráktól. Ehelyett a hangsúly az információs architektúra új megközelítéseinek kialakítására helyeződik át, ami a vállalatokat modern adatvezérelt vállalkozásokká alakítja át. Előbb-utóbb maga az építőipar is meg fogja tenni ezt a lépést, és a fokozatos digitális evolúcióból egy valódi, minden vállalatot érintő digitális forradalom fog átmenni.

Az adatvezérelt üzleti folyamatokra való áttérés nem lesz könnyű. Sok vállalatnak kihívásokkal kell szembenéznie, mivel a vezetők nem mindig értik, hogyan használhatják fel a kaotikus adathalmazokat a hatékonyság és az üzleti növekedés javítására.

Ez a könyv az adatok világába kalauzol el, ahol az információ az üzleti folyamatok hatékonyságát és fenntarthatóságát meghatározó kulcsfontosságú stratégiai erőforrássá válik. Az információk gyors növekedésével a vállalatok új kihívásokkal szembesülnek. A digitális átalakulás már nem csak egy divatszó - szükségszerűsséggé válik.



4. ábra Az adatok és a folyamatok az építés alapját képezik.

Az átalakulás megértése azt jelenti, hogy egyszerű szavakkal meg tudjuk magyarázni az összetett dolgokat. Ezért a könyv közérthető nyelven íródott, és a szerző illusztrációi kísérik, amelyeket kifejezetten azért készített, hogy a kulcsfogalmakat világosan elmagyarázza. Ezek az ábrák, grafikonok és vizualizációk úgy vannak kialakítva, hogy elhárítsák az érzékelés akadályait, és még azok számára is érhetővé tegyék az anyagot, akik korábban túl bonyolultnak tartották az ilyen témaikat. A könyvben található összes illusztrációt, diagramot és grafikát a szerző készítette, és kifejezetten a szövegben leírt kulcsfogalmak vizualizálására terveztek.

Egy kép többet ér ezer szónál [3].

- Fred R. Barnard, angol illusztrátor, 1927.

Az elmélet és a gyakorlat összekapcsolása érdekében mesterséges intelligencia eszközöket (különösen nyelvi modellek) fogunk használni, amelyek lehetővé teszik, hogy mély programozási ismeretek nélkül fejlesszen ki megoldásokat. Ha a gyakorlati anyag felé orientálódik, és jobban érdekli az adatokkal való gyakorlati munka, akkor az első bevezető részt átugorhatja, és rögtön a könyv második részére térhet át, ahol a konkrét példák és esetek ismertetése kezdődik.

Ne támasszon azonban túlzott elvárásokat az AI (mesterséges intelligencia), a gépi tanulás és az LLM (Large Language Models) eszközökkel szemben. Minőségi bemeneti adatok és a téma mély megértése nélkül még a legfejlettebb algoritmusok sem tudnak megbízható és értelmes eredményeket szolgáltatni.

Satya Nadella, a Microsoft vezérigazgatója 2025 elején a mesterséges intelligencia buborékveszélyére figyelmeztet [4], a jelenlegi hype-ot a dot-com buborékhoz hasonlítva. Hangsúlyozza, hogy az AGI (Artificial General Intelligence) mérföldkövek elérésére vonatkozó állítások megfelelő indoklás nélkül "értelmetlen manipuláció a mérőszámokkal". Nadella úgy véli, hogy az AI valódi sikerét a globális GDP-növekedéshez való hozzájárulásán kell mérni, nem pedig a divatszavak túlzott hangsúlyozásán.

Az új technológiákról és koncepciókról szóló összes divatos szó mögött összetett és fáradságos munka rejlik az adatminőség biztosítása érdekében, az üzleti folyamatok paraméterezése és az eszközök valós feladatokhoz való igazítása.

Az adatvezérelt megközelítés nem olyan termék, amelyet csak úgy letölthet vagy megvásárolhat. Ez egy stratégia, amelyet ki kell építeni. A meglévő folyamatok és problémák újragondolásával kezdődik, majd fegyelmezett mozgást igényel a választott irányba.

A vezető szoftverfejlesztők és alkalmazásgyártók nem lesznek a változás motorjai az építőiparban, sokuk számára az adatvezérelt megközelítés veszélyt jelent a bevett üzleti modelljükre.

Más iparágak [az építőiparral ellentétben], például az autóipar, már átmentek radikális és bomlásztó változásokon, és a digitális átalakulásuk már javában zajlik. Az építőipari vállalatoknak gyorsan és határozottan kell cselekedniük: a fürge cégek hatalmas hasznat fognak húzni, míg azok, akik haboznak, komoly kockázatot vállalnak. Gondoljunk csak a digitális fényképezés által ebben az iparágban okozott felfordulásra [5].

- A Világgazdasági Fórum jelentése: Az építőipar jövőjének alakítása, 2016.

Azok a vállalatok, amelyek időben felismerik az új megközelítés lehetőségeit és előnyeit, fenntartható versenyelőnyre tesznek szert, és képesek lesznek fejlődni és növekedni anélkül, hogy a nagy gyártók megoldásaitól függenének.

Itt az alkalom, hogy ne csak átvészélje az információ digitalizálásának közelgő viharát, hanem át is vegye annak irányítását. Ebben a könyvben nem csupán az iparág jelenlegi helyzetének elemzését találja, hanem konkrét ajánlásokat is, amelyekkel újragondolhatja és átalakíthatja folyamatait és vállalkozását, hogy vezető szerepet tölthessen be az építőipar új korszakában, és növelje szakmai tapasztalatát.

Az építőipar digitális jövője nem csupán az új technológiák és programok használatáról szól, hanem az adatkezelés és az üzleti modellek alapvető újragondolásáról.

Készen áll az Ön vállalata erre a stratégiai változásra?

TARTALOMJEGYZÉK

BEVEZETÉS	1
TARTALOMJEGYZÉK.....	I
II RÉSZ AZ AGYAGTÁBLÁKTÓL A DIGITÁLIS FORRADALOMIG: HOGYAN FEjlŐDött AZ INFORMÁCIÓ AZ ÉPÍTŐIPARBAN	2
FEJEZET 2.1. AZ ADATFELHASZNÁLÁS FEjlŐDÉSE AZ ÉPÍTŐIPARBAN	3
Az adatkorszak születése az építőiparban	3
Az agyagtól és a papirusztól a digitális technológiáig	4
A folyamat mint az adatvezérelt élmény eszköze.....	5
Az építési folyamatokra vonatkozó információk digitalizálása	7
FEJEZET 2.2. TECHNOLÓGIÁK ÉS IRÁNYÍTÁSI RENDSZEREK A MODERN ÉPÍTŐIPARBAN	11
A digitális forradalom és a moduláris MRP/ERP -rendszerek megjelenése.....	11
Adatkezelési rendszerek: az adatbányászattól az üzleti kihívásokig.....	13
Vállalati micélium: hogyan kapcsolódnak az adatok az üzleti folyamatokhoz	16
FEJEZET 2.3. A DIGITÁLIS FORRADALOM ÉS AZ ADATROBBANÁS	20
Az adatboom kezdete, mint evolúciós hullám	20
Egy modern vállalatnál keletkező adatmennyiség.....	22
Az adattárolás költségei: gazdasági szempontok.....	23
Az adathalmozás határai: a tömegtől a jelentésig	25
Következő lépések: az adatelmélettől a gyakorlati változásig.....	27
III RÉSZ HOGYAN FULLAD BELE AZ ÉPÍTŐIPAR AZ ADATKÁOSZBA.....	28
FEJEZET 3.1. AZ ADATOK SZÉTTÖREDEZETTSÉGE ÉS SILÓK	29
Minél több eszköz, annál hatékonyabb a vállalkozás?	29
Adatsilók és azok hatása a vállalati teljesítményre.....	31
Duplikáció és az adatok minőségének hiánya a széttagoltság következményeként.	34
HiPPO vagy a vélemények veszélye a döntéshozatalban.....	36
Az üzleti folyamatok összetettségének és dinamizmusának folyamatos növekedése	39
A negyedik ipari forradalom (Ipar 4.0) és az ötödik ipari forradalom (Ipar 5.0) az építőiparban.....	42
FEJEZET 3.2. A KÁOSZ RENDDÉ ALAKÍTÁSA ÉS A KOMPLEXITÁS CSÖKKENTÉSE	45
A redundáns kód és a zárt rendszerek mint a termelékenység javításának akadályai.....	45
A silóktól az egységes adattárházig.....	47
Az integrált tárolórendszerek lehetővé teszik az átállást az AI ügynökökre.....	48

Az adatgyűjtéstől a döntéshozatalig: az automatizáláshoz vezető út	51
Következő lépések: a káoszból kezelhető rendszerré alakítása.....	53
IV RÉSZ ADATKERET AZ ÉPÍTŐIPARI ÜZLETI FOLYAMATOKBAN	55
FEJEZET 4.1. ADATTÍPUSOK AZ ÉPÍTŐIPARBAN.....	56
A legfontosabb adattípusok az építőiparban.....	56
Strukturált adatok	60
Relációs adatbázisok RDBMS és az SQL lekérdezési nyelv	61
SQL - lekérdezések az adatbázisokban és új trendek.....	64
Strukturálatlan adatok	66
Szöveges adatok: a strukturálatlan káosz és a strukturált káosz között y	67
Félíg strukturált és lazán strukturált adatok	68
Geometriai adatok és alkalmazása.....	69
CAD adatok: a tervezéstől az adattárolásig.....	72
A BIM (BOM) koncepció megjelenése és a CAD felhasználása a folyamatokban.....	75
FEJEZET 4.2. ADATHARMONIZÁCIÓ ÉS STRUKTURÁLÁS.....	81
Rendszerek adatokkal való feltöltése az építőiparban	81
Adattranszformáció: a modern üzleti elemzés kritikus alapjai	83
Adatmodellek: kapcsolatok az adatokban és az elemek közötti kapcsolatok	87
Szabadalmaztatott formátumok és hatásuk a digitális folyamatokra	91
A nyílt formátumok megváltoztatják a digitalizáláshoz való hozzáállást.....	95
Paradigmaváltás: A nyílt forráskódú mint a szoftvergyártók dominanciájának korszakának vége	96
Strukturált nyílt adatok: a digitális átalakulás alapja	99
FEJEZET 4.3. LLM ÉS SZEREPÜK AZ ADATFELDOLGOZÁSBAN ÉS AZ ÜZLETI FOLYAMATOKBAN.....	102
LLM csevegőszobák: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok az adatfeldolgozási folyamatok automatizálására.....	102
Nagy LLM nyelvi modellek: hogyan működik?	103
Helyi LLM-ek használata az érzékeny vállalati adatok esetében	106
Az AI teljes ellenőrzése a vállalaton belül és a saját LLM bevetésének módja	108
RAG: Intelligens LLM - a vállalati adatokhoz hozzáférő asszisztensek	110
FEJEZET 4.4. IDE AZ LLM TÁMOGATÁSSAL ÉS JÖVŐBELI PROGRAMOZÁSI VÁLTOZÁSOKKAL.....	112
Az IDE kiválasztása: az LLM kísérletekből az üzleti megoldásokig	112
IDE LLM-támogatással és jövőbeli programozási változtatások	114
Python Pandas: nélkülözhetetlen eszköz az adatokkal való munkához.	115
DataFrame: univerzális táblázatos adatformátum.....	119

Következő lépések: fenntartható adatkeret kialakítása	122
V RÉSZ ADATMINŐSÉG: SZERVEZÉS, STRUKTURÁLÁS, MODELLEZÉS	124
FEJEZET 5.1. AZ ADATOK STRUKTURÁLT FORMÁBA TÖRTÉNŐ ÁTALAKÍTÁSA	125
Dokumentumok, PDF, képek és szövegek strukturált formátummá alakításának megtanulása.	125
Példa egy PDF -dокументum táblázattá alakítására	126
JPEG, PNG kép átalakítása strukturált formába.....	130
Szöveges adatok strukturált formába történő átalakítása	133
A CAD-adatok (BIM) strukturált formába történő konvertálása.....	136
A CAD megoldások szállítói a strukturált adatok felé mozdulnak el.....	141
FEJEZET 5.2. OSZTÁLYOZÁS ÉS INTEGRÁCIÓ: AZ ÉPÍTÉSI ADATOK KÖZÖS NYELVE.....	144
A döntéshozatal sebessége az adatok minőségétől függ	144
Adatszabványosítás és -integráció.....	145
A digitális átjárhatóság a követelményekkel kezdődik	148
Az építés közös nyelve: az osztályozók szerepe a digitális átalakulásban	151
Masterformat, OmniClass, Uniclass és CoClass: az osztályozási rendszerek fejlődése	153
FEJEZET 5.3. ADATMODELLEZÉS ÉS KIVÁLÓSÁGI KÖZPONT	158
Adatmodellezés: fogalmi, logikai és fizikai modell	158
Gyakorlati adatmodellezés az építőiparban	162
Adatbázis létrehozása az LLM segítségével	164
Adatmodellezési kiválósági központ (CoE).....	166
FEJEZET 5.4. A KÖVETELMÉNYEK RENDSZEREZÉSE ÉS AZ INFORMÁCIÓK VALIDÁLÁSA	169
A követelmények összegyűjtése és elemzése: A kommunikáció strukturált adatokká történő átalakítása.....	169
Folyamatábrák és a fogalmi keretek hatékonysága.....	173
Strukturált követelmények és RegEx reguláris kifejezések.....	175
Adatgyűjtés az ellenőrzési folyamathoz.....	180
A adatok ellenőrzése és az ellenőrzés eredményei	182
A ellenőrzési eredmények vizualizálása	187
Az adatminőség-ellenőrzések összehasonlítása az emberi élet szükségleteivel	189
Következő lépések: az adatok pontos számításokká és tervkké alakítása	191
VI RÉSZ KÖLTSÉG- ÉS IDŐSZÁMÍTÁS: AZ ADATOK BEÉPÍTÉSE AZ ÉPÍTÉSI FOLYAMATOKBA	193
FEJEZET 6.1. ÉPÍTÉSI PROJEKTEK KÖLTSÉGSZÁMÍTÁSAI ÉS BECSLÉSEI	194
Építőipari alapismeretek: mennyiségi-, költségi- és időbecslés	194
A projektek becsült költségeinek kiszámítási módszerei	195

Erőforrás-alapú módszer költségszámítás és becslés az építőiparban	196
Építőipari erőforrások adatbázisa: építőanyagok és építési munkák katalógusa	196
Számítások összeállítása és a munkák költségszámítása az erőforrásbázis alapján.	198
A projekt végső költségszámítása: a becsléstől a költségvetésig.....	203
FEJEZET 6.2. MENNYISÉGI SZÁMBAVÉTEL ÉS AUTOMATIKUS BECSLÉSEK ÉS ÜTEMTERVEK KÉSZÍTÉSE	207
A 3D-ről a 4D-re és 5D-re való áttérés: volumetrikus és mennyiségi paraméterek használata	207
5D attribútumok és az attribútum térfogatának kinyerése a CADról	207
QTO Quantity Take-Off: projektadatok csoportosítása attribútumok szerint	211
QTO automatizálás az LLM és a strukturált adatok használatával.....	216
QTO A teljes projekt számítása Excel táblázatból származó csoportszabályok segítségével.	220
FEJEZET 6.3. 4D, 6D -8D ÉS A SZÉNDIOXID-KIBOCSÁTÁS KISZÁMÍTÁSA CO₂.....	226
4D modell: az idő beépítése az építési becslésekbe	226
Építési ütemterv és annak automatizálása a költségszámítási adatok alapján	227
Kibővített attribútumrétegek 6D -8D: az energiahatékonyságtól a biztonság garantálásáig.....	229
A CO ₂ becslése és az építési projektek szén-dioxid-kibocsátásának kiszámítása	232
FEJEZET 6.4. ÉPÍTŐIPARI ERP ÉS PMIS RENDSZEREK.....	237
Építőipari ERP -rendszerek a számítások és becslések példáján.....	237
PMIS: Közvetítő az ERP és az építkezés között.....	242
Spekuláció, profit, elszigeteltség és a hiányos átláthatóság az ERP és a PMIS rendszerekben	243
A zárt ERP /PMIS korszakának vége: az építőiparnak új megközelítésekre van szüksége.....	246
Következő lépések: a projektadatok hatékony felhasználása.....	247
VII RÉSZ CAD ÉS BIM: MARKETING, VALÓSÁG ÉS A TERVEZÉSI ADATOK JÖVŐJE AZ ÉPÍTŐIPARBAN	
252	
FEJEZET 7.1. A BIM- KONCEPCIÓK MEGJELENÉSE AZ ÉPÍTŐIPARBAN.....	253
A BIM és a nyílt BIM mint a CAD- gyártók marketingkoncepcióinak kialakulásának története.	253
A BIM valósága: integrált adatbázisok helyett - zárt moduláris rendszerek.....	256
A nyílt formátumú IFC megjelenése az építőiparban.....	258
IFC formátum probléma a geometriai magtól függően	260
Megjelenés a szemantika és az ontológia téma körének felépítésében	263
Miért nem váltják be a szemantikus technológiák az építőiparban a hozzájuk fűzött reményeket? 265	
FEJEZET 7.2. ZÁRT PROJEKTFORMÁTUMOK ÉS INTEROPERABILITÁSI PROBLÉMÁK.....	269
Zárt adatok és csökkenő termelékenység: a CAD-ipar zsákutcája (BIM)	269
Az interoperabilitás mítosza a CAD-rendszerek között.....	271

Tovább az USD-hez és a szemcsés adatokhoz m	275
FEJEZET 7.3. GEOMETRIA AZ ÉPÍTŐIPARBAN: A VONALAKTÓL A KÖBMÉTEREKIG.....	279
Amikor a vonalak pénzzé válnak, vagy hogy miért van szükségük az építőknek geometriára	279
Vonalakból kötetekké: Hogyan lesz a terület és a térfogat adat.....	279
Mozgás a MESH, USD és poligonok felé: a tesszelláció használata a geometriához	281
LOD, LOI, LOMD - a részletek egyedi osztályozása a CAD-ben (BIM).....	283
Új CAD-szabványok (BIM) - AIA, BEP, IDS, LOD, COBie	286
FEJEZET 7.4. TERVEZÉSI PARAMÉTEREZÉS ÉS AZ LLM HASZNÁLATA A CAD ÜZEMELTETÉSHEZ	291
A CAD-adatok egyediségének illúziója (BIM): az analitika és a nyílt formátumok felé vezető út.....	291
Tervezés paramétereken keresztül: a CAD és a BIM jövője.....	294
Az LLM megjelenése a tervezési CAD adatfeldolgozási folyamatokban.....	297
DWG -fájlok automatizált elemzése az LLM és Pandas segítségével.....	300
Következő lépések: a zárt formátumokról a nyílt adatokra való áttérés.....	306
VIII RÉSZ ADATVEZÉRELT DÖNTÉSHOZATAL, ANALITIKA, AUTOMATIZÁLÁS ÉS GÉPI TANULÁS	308
FEJEZET 8.1. ADATELEMZÉS ÉS ADATVEZÉRELT DÖNTÉSHOZATAL.....	309
Az adatok mint a döntéshozatalban használt erőforrás	309
A adatok vizualizálása: a megértés és a döntéshozatal kulcsa	312
KPI-k és ROI	314
Műszerfalak és műszerfalak: a mérőszámok vizualizálása a hatékony irányítás érdekében.....	316
Adatelemzés és a kérdésfeltevés művészete.....	318
FEJEZET 8.2. ADATÁRAMLÁS KÉZI MUNKA NÉLKÜL: MIÉRT VAN SZÜKSÉG AZ ETL-RE	321
ETL automatizálás: költségcsökkentés és gyorsabb adatkezelés.....	321
ETL Kivonat: adatgyűjtés	325
ETL Transform: validálási és transzformációs szabályok alkalmazása	328
ETL Load: A eredmények grafikonokon és diagramokon való megjelenítése.	330
ETL Load: PDF dokumentumok automatikus létrehozása	335
ETL Load: dokumentumok automatikus generálása FPDF-ből	336
ETL Load: Jelentések készítése és betöltése más rendszerekbe.....	340
ETL az LLM-mel: adatok vizualizálása PDF -dокументumokból	341
FEJEZET 8.3. AUTOMATIKUS ETL SZÁLLÍTÓSZALAG (CSŐVEZETÉK)	346
Pipeline: Automatikus ETL szállítószalag adatok.....	346
Csővezeték -ETL adatérvényesítési folyamat LLM-mel	350
Pipeline -ETL: a projektelemek adatainak és információinak ellenőrzése a CAD-ben (BIM).	352
FEJEZET 8.4. AZ ETL ÉS MUNKAFOLYAMATOK ÖSSZEHANGOLÁSA: GYAKORLATI MEGOLDÁSOK	358

DAG és Apache Airflow: munkafolyamatok automatizálása	358
Apache Airflow: gyakorlati alkalmazás az ETL automatizálására	359
Apache NiFi útválasztáshoz és adatkonverzióhoz.....	363
n8n Alacsony kódú, No-Code folyamat-orchestrálás.....	364
Következő lépések: a manuális műveletekről az analitikán alapuló megoldásokra való áttérés	367
IX RÉSZ ADATTÁROLÁS ÉS -KEZELÉS AZ ÉPÍTŐIPARBAN369	
FEJEZET 9.1. ADATINFRASTRUKTÚRA: A TÁROLÁSI FORMÁTUMOKTÓL A DIGITÁLIS ADATTÁRAKIG 370	
Adatatomok: a hatékony információkezelés alapja.....	370
Információ tárolása: fájlok vagy adatok.....	371
Nagy adattárolás: a népszerű formátumok és hatékonyságuk elemzése.....	373
Tárolás optimalizálása az Apache Parquet segítségével	376
DWH: Adattárház adattárházak	378
Data Lake - az ETL evolúciója az ELT: a hagyományos tisztítástól a rugalmas feldolgozásig	380
Data Lakehouse architektúra: a raktárak és adattávok szinergiája.....	382
CDE, PMIS, ERP vagy DWH és Data Lake	384
FEJEZET 9.2. ADATTÁRHÁZ-KEZELÉS ÉS KÁOSZMEGELŐZÉS.....	388
Vektoradatbázisok és a határoló doboz.....	388
Adatkezelés), adatminimalizmus) és adatmocsár).....	391
DataOps és VectorOps: új adatszabványok.....	394
A következő lépések: a kaotikus tárolástól a strukturált tárolásig	396
X RÉSZ BIG DATA, GÉPI TANULÁS ÉS ELŐREJELZÉSEK398	
FEJEZET 10.1. BIG DATA ÉS ANNAK ELEMZÉSE	399
Big data az építőiparban: az intuíciótól a kiszámíthatóságig	399
A nagyméretű adatok megvalósíthatóságának megkérdőjelezése: korreláció, statisztika és adatmintavételezés	400
Nagy adatok: San Francisco egymillió építési engedélyének adatainak elemzése	403
Példa a CAD-adatokon alapuló nagyméretű adatokra (BIM)	409
IoT A dolgok internete és az intelligens szerződések.....	413
FEJEZET 10.2. GÉPI TANULÁS ÉS ELŐREJELZÉSEK	417
A gépi tanulás és a mesterséges intelligencia meg fogja változtatni az építkezésünket.....	417
A szubjektív értékeléstől a statisztikai előrejelzésig	419
Titáni adatkészlet: Hello World az analitika világában adatok és nagy adatok	421
Gépi tanulás akcióban: a Titanic utasaitól a projektmenedzsmentig	426
Előrejelzések és előrejelzések történeti adatok alapján.....	430

A gépi tanulás kulcsfogalmai.....	432
FEJEZET 10.3. A KÖLTSÉGEK ÉS HATÁRIDŐK ELŐREJELZÉSE GÉPI TANULÁSSAL.....	435
Egy példa a gépi tanulás használatára a projektköltségek és határidők meghatározására.....	435
Projektköltség és -idő előrejelzése lineáris regresszió segítségével	437
A projektköltségek és -idők előrejelzése a K-közelebbi szomszéd algoritmus segítségével (k-NN) ..	440
Következő lépések: a tárolástól az elemzésig és előrejelzésig.....	444
XI RÉSZ AZ ÉPÍTŐIPAR A DIGITÁLIS KORBAN. LEHETŐSÉGEK ÉS KIHÍVÁSOK	447
FEJEZET 11.1. TÚLÉLÉSI STRATÉGIÁK: VERSENYELŐNYÖK KIÉPÍTÉSE	448
Korrelációk helyett számítások: az építőipari analitika jövője.....	448
Adatvezérelt megközelítés az építőiparban: az infrastruktúra új szintje.....	451
A következő generációs digitális iroda: hogyan változtatja meg az AI a munkaterületet?	453
A nyílt adatok és az uberizáció veszélyét jelent a meglévő építőipari vállalkozásokra nézve	455
Az uberizáció megoldatlan problémái mint utolsó esély az idő kihasználására az átalakulás érdekében	458
FEJEZET 11.2. GYAKORLATI ÚTMUTATÓ AZ ADATVEZÉRELT MEGKÖZELÍTÉS MEGVALÓSÍTÁSHOZ .	463
Az elmélettől a gyakorlatig: Útiterv az építőipar digitális átalakulásához	463
A digitális alapok megteremtése: 1-5 lépés a digitális érettség felé	465
Az adatokban rejlő lehetőségek felszabadítása: 5-10 lépés a digitális érettség felé.....	469
Átalakítási ütemterv: a káosztól az adatvezérelt vállalatig.....	476
Építkezés az ipar 5.0-ban: Hogyan lehet pénzt keresni, ha már nem lehet elbújni.....	479
KÖVETKEZTETÉS	481
A SZERZŐRŐL	484
TOVÁBBI KAPCSOLAT	485
FORDÍTÁS KOMMENTÁR.....	485
EGYÉB KÉSZSÉGEK ÉS FOGALMAK	486
FOGALOMTÁR	490
HIVATKOZÁSOK ÉS ONLINE ANYAGOK LISTÁJA	497
TÁRGYI INDEX	513

MAXIMÁLIS KÉNYELEM A NYOMTATOTT VERZIÓVAL

Ön birtokában van a **Data-Driven Construction** ingyenes digitális verziója. A kényelmesebb munkavégzés és az anyagokhoz való gyors hozzáférés érdekében javasoljuk, hogy figyeljen a nyomtatott kiadásra:



■ **Mindig kéznél van:** a nyomtatott formátumú könyv megbízható munkaeszközzé válik, amely lehetővé teszi, hogy bármilyen munkaszituációban gyorsan megtalálja és használja a megfelelő ábrákat és diagramokat.

■ **Kiváló minőségű illusztrációk:** a nyomtatott kiadásban szereplő összes kép és grafika maximális minőségen jelenik meg.

■ **Gyors hozzáférés az információkhoz:** könnyű navigáció, jegyzetek, könyvjelzők készítése és a könyvvel való munka bárhol.

A könyv teljes nyomtatott változatának megvásárlásával egy kényelmes eszközt kap az információkkal való kényelmes és hatékony munkavégzéshez: a vizuális anyagok gyors használatának lehetőségét a minden nap feladatakban, a szükséges sémák gyors megtalálását és jegyzetelését. Emellett vásárlásával a nyílt tudás terjesztését is támogatja.

Rendelje meg a könyv nyomtatott változatát a következő címen:
datadrivenconstruction.io/books



I RÉSZ

AZ AGYAGTÁBLÁKTÓL A DIGITÁLIS FORRADALOMIG: HOGYAN FEJLŐDÖTT AZ INFORMÁCIÓ AZ ÉPÍTŐIPARBAN

A könyv első része az építőipari adatkezelés történelmi fejlődését vizsgálja, a fizikai adathordozókon tárolt kezdetleges nyilvántartásoktól a modern digitális ökoszisztemákig. Elemzi az információkezelési technológiák átalakulását, az ERP - rendszerek megjelenését és az adatok széttöredezettségének az üzleti folyamatok hatékonyságára gyakorolt hatását. Különös figyelmet fordít az információ digitalizálásának folyamatára és a szubjektív szakértői megítélés helyett az objektív elemzés növekvő jelentőségére. Részletesen megvizsgálják az információmennyiségek exponenciális növekedését, amellyel a modern építőiparnak szembe kell néznie, és a vállalati rendszerekkel kapcsolatos kihívásokat. A negyedik és ötödik ipari forradalom összefüggésében vizsgálják az építőipar helyzetét, valamint a mesterséges intelligencia és az adatközpontú megközelítések használatának lehetőségeit a fenntartható versenyelőny megteremtése érdekében.

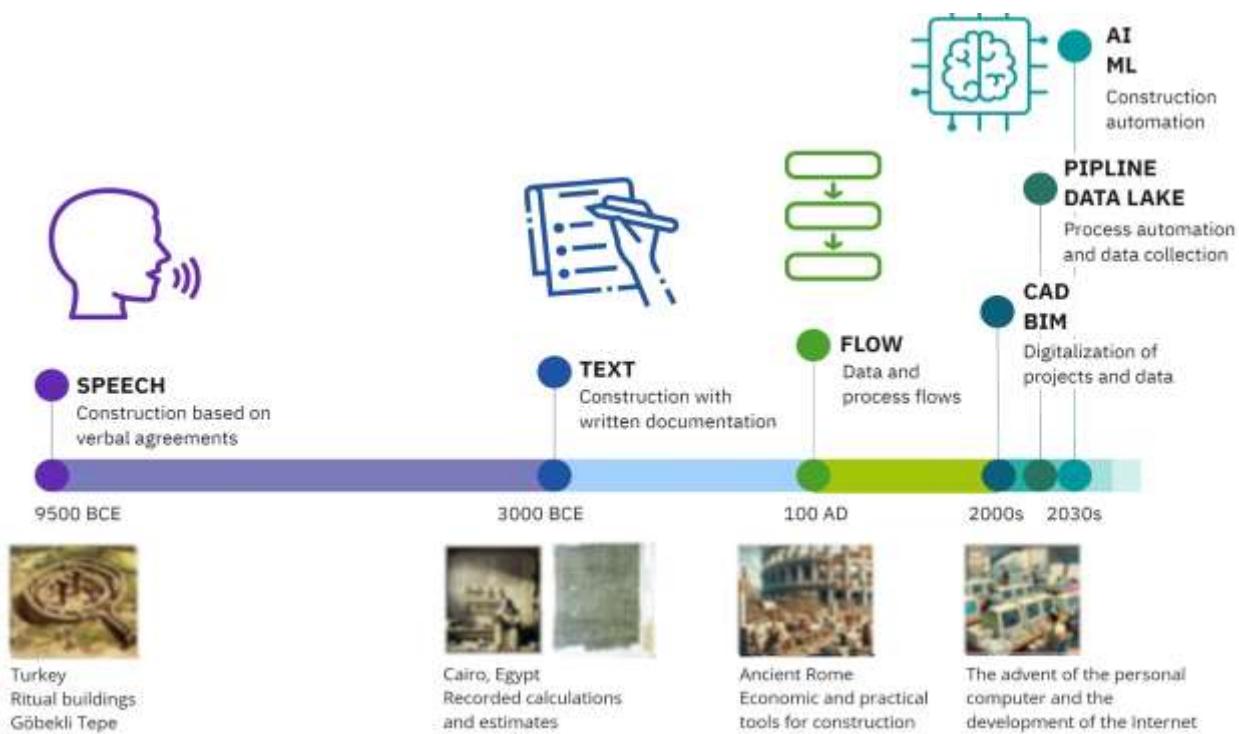
FEJEZET 1.1.

AZ ADATFELHASZNÁLÁS FEJLŐDÉSE AZ ÉPÍTŐIPARBAN

Az adatkorszak születése az építőiparban

Körülbelül 10 000 évvel ezelőtt, a neolitikumban az emberiség forradalmi változást hajtott végre fejlődésében: a nomád életmódot felhagyta a letelepedett életmód javára, ami az első primitív, agyagból, fából és kőből készült épületek megjelenéséhez vezetett [6]. Ettől a pillanattól kezdődik az építőipar története.

A civilizációk fejlődésével az építészet egyre összetettebbé vált, és ez vezetett az első rituális templomokhoz és középületekhez. Az építészeti tervek növekvő bonyolultsága miatt az ókor mérnökeinek és menedzsereinek meg kellett készíteniük az első feljegyzéseket és számításokat. Az agyagtáblákon és papirouszokon található első feljegyzések gyakran tartalmazták a szükséges építőanyagok mennyiségének és költségének kiszámítását, valamint az elvégzett munkáért járó fizetés kiszámítása mögötti logika leírását [7]. Így kezdődött az adatok építőipari felhasználásának korszaka - jóval a modern digitális technológiák megjelenése előtt (1.1-1. ábra).

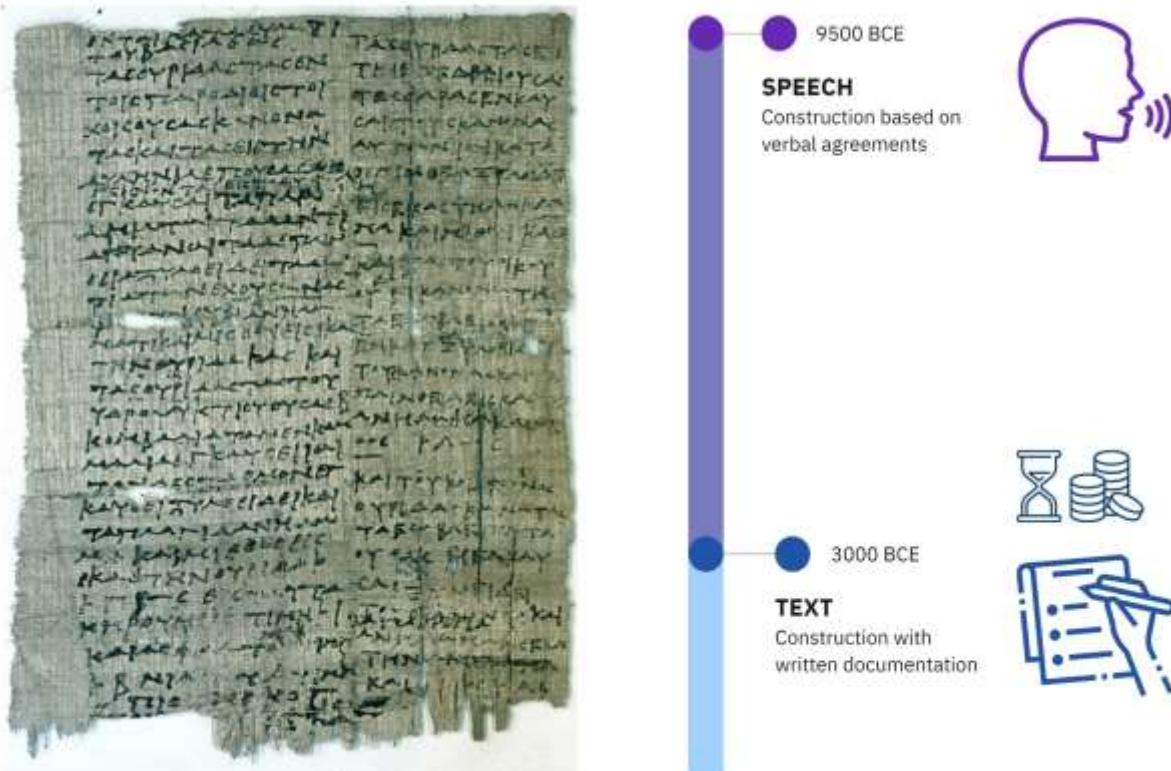


Ábra: 1.1-1 Az építőipari informatika fejlődésének kronológiája: a szóbeli információtól a mesterséges intelligenciáig.

Az agyagtól és a papirusztól a digitális technológiáig

Az első dokumentált bizonyítékok az építésben a piramisépítés időszakára, Kr. e. 3000-4000 körüli időszakra nyúlnak vissza[7]. Azóta az írásos feljegyzések vezetése megkönnyítette és kísérte az építőipar fejlődését, lehetővé téve a tudás felhalmozását és rendszerezését, ami a következő 10 000 év során jelentős újításokhoz vezetett az építési módszerek és az építészet terén.

Az első fizikai építőipari adathordozók, mint például az évezredekkel ezelőtti agyagtáblák, papiruszok (1.1-2. ábra) vagy az 1980-as években az "A0" papír adatrögzítésére történő felhasználása eredetileg nem arra irányult, hogy ezeket az információkat új projekteknél alkalmazzák. Az ilyen feljegyzések fő célja a projekt aktuális állapotának részletezése volt, beleértve a szükséges anyagok és a munkálatok költségeinek kiszámítását. Hasonlóképpen, a mai világban a digitális tervezési adatok és modellek rendelkezésre állása nem minden garantálja azok alkalmazását a jövőbeli projektekben, és gyakran elsősorban a szükséges anyagok és építési költségek aktuális számításainak információjaként szolgál.



1.1-2. ábra Egy i. e. 3. századi papirusz, amely egy királyi palota különböző típusú ablakainak enkausztikus technikával történő festésének költségeit írja le.

Az emberiségnek körülbelül 5000 évébe telt, hogy az építési projektek irányításában a szóbeli beszélgetésekről áttérjen az írásos dokumentumokra, és ugyanennyi időbe telt, hogy a tervezés és ellenőrzés elsődleges forrásaként a papírról a digitális adatokra térjen át.

Ahogyan a kereskedelmi és monetáris kapcsolatok fejlődése ösztönözte az írás és az első jogászok megjelenését a viták megoldására, úgy az építőiparban az anyagköltségek és a munkamennyiségek első nyilvántartása vezetett az első menedzserek megjelenéséhez az építőiparban, akiknek feladatai közé tartozott a projektek ütemezésére és költségeire vonatkozó kulcsfontosságú információk dokumentálása, nyomon követése és felelőssége.

Ma már az adatok sokkal jelentősebb szerepet játszanak: nemcsak a meghozott döntéseket rögzítik, hanem a jövő előrejelzésének és modellezésének eszközévé is válnak. Ez az az alap, amelyre a projektmenedzsment modern folyamatszemlélete épül - a felhalmozott tapasztalatokat strukturált és ellenőrizhető adatokon alapuló döntéshozatali rendszerré alakítja.

A folyamat mint az adatvezérelt élmény eszköze

Minden folyamat középpontjában a múltbeli tapasztalatoknak a jövő tervezésének eszközévé történő átalakítása áll. A modern értelemben vett tapasztalat olyan strukturált adathalmaz, amelyet elemezve megalapozott előrejelzéseket lehet készíteni.

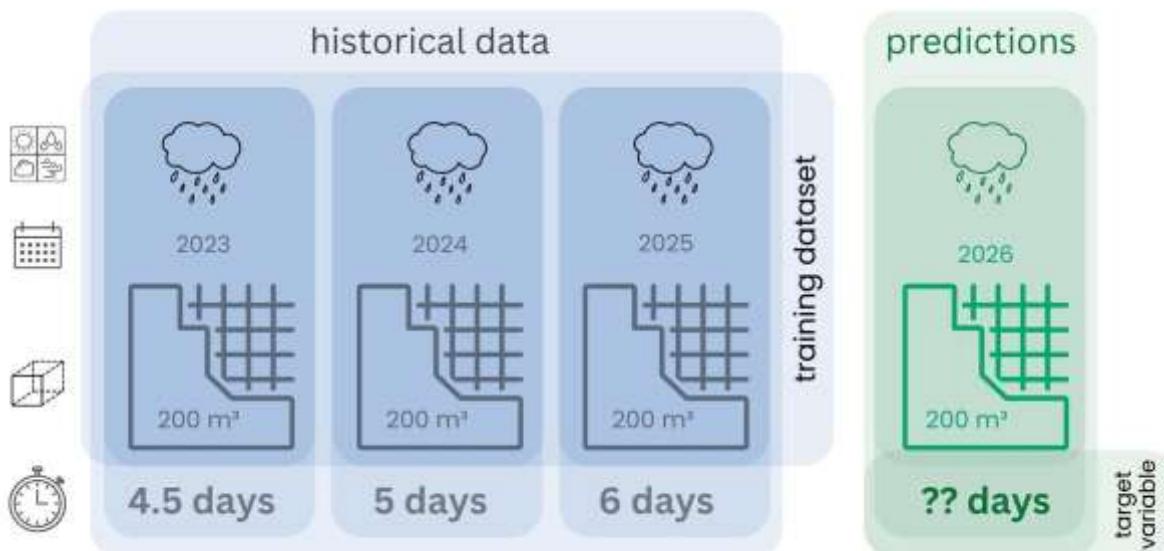
Az előrejelzés alapjául a múltbeli adatok szolgálnak, mivel egyértelműen bemutatják az elvégzett munka eredményeit, és betekintést nyújtanak az ezeket az eredményeket befolyásoló tényezőkbe.

Vegyük egy konkrét példát a monolitikus építkezésből: általában a munkálatok ütemezésének tervezésekor figyelembe veszik a beton mennyiségét, a szerkezet összetettségét és az időjárási körülményeket. Tegyük fel, hogy egy adott építésvezető vagy a vállalat elmúlt három év (2023-2025) történeti adatai azt mutatják, hogy egy 200 m²-es monolit szerkezet kiöntése esős időben 4,5-6 napot vett igénybe (1.1-3. ábra). Ezek a felhalmozott statisztikák képezik az alapját az átfutási idők előrejelzésének és az erőforrások kalkulációjának, amikor a jövőbeli projektek hasonló munkáit tervezik. E múltbeli adatok alapján a művezető vagy a becslő a tapasztalatok alapján megalapozott előrejelzést tud készíteni a 2026-ban hasonló körülmények között a jövőbeni hasonló munkák elvégzéséhez szükséges időre vonatkozóan.

Az időelemzés esetében az analitikai folyamat olyan mechanizmusként működik, amely a széttagolt adatokat strukturált tapasztalatokká, majd pontos tervezési eszközzé alakítja. Az adatok és a folyamatok egyetlen ökoszisztemát alkotnak, ahol az egyik nem létezhet a másik nélkül.

Számold meg, ami megszámolható, mérd meg, ami mérhető, és tudd mérhetővé, ami nem mérhető [8].

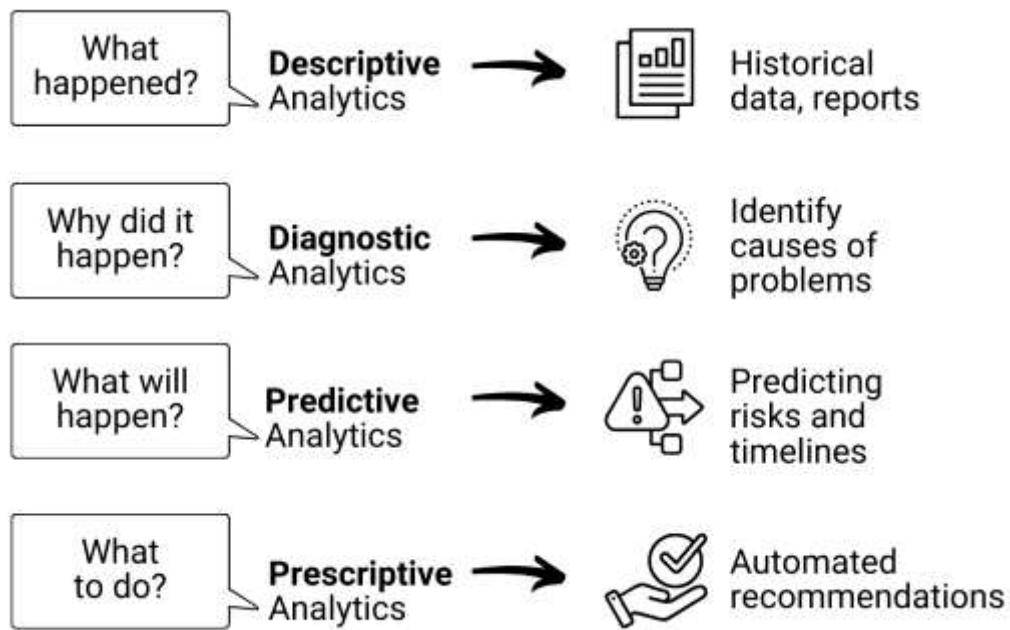
- Galileo Galilei



1.1-3. ábra A múltbeli adatok képzési adatkészletként szolgálnak a jövőbeli értékek egyikének előrejelzéséhez.

A mai üzleti környezetben az adatalemzés a hatékony projektmenedzsment, a folyamatoptimalizálás és a stratégiai döntéshozatal kritikus elemévé válik. Az építőipar fokozatosan elsajátítja az analitika négy kulcsfontosságú szintjét, amelyek mindegyike egy-egy konkrét kérdésre ad választ és egyedi előnyöket biztosít (1.1-4. ábra):

- **Leíró analitika** - választ ad a "mi történt?" kérdésre, és történeti adatokat és jelentéseket nyújt a múltbeli eseményekről és eredményekről: az elmúlt három évben (2023-2025) esős időben 4,5 és 6 nap között volt egy 200 m²-es monolit szerkezet kiöntése.
- **Diagnosztikai analitika** - a problémák okainak azonosításával válaszol a "miért történt ez?" kérdésre: az elemzés azt mutatja, hogy a monolit szerkezet öntési ideje megnőtt az esős időjárás miatt, ami lelassította a beton száradási folyamatát.
- **Előrejelző analitika** - előrettekintő, a lehetséges kockázatok és átfutási idők előrejelzése a "mi fog történni?" kérdés megválaszolásával: a múltbeli adatok alapján azt jósolják, hogy egy hasonló, 200 m²-es monolit szerkezet kiöntése esős időben 2026-ban körülbelül 5,5 napot vesz igénybe, figyelembe véve minden ismert tényezőt és tendenciát.
- **Prescriptive analytics** - automatizált ajánlásokat ad, és választ ad a "mit tegyek?" kérdésre, lehetővé téve a vállalatok számára, hogy az optimális lépéseket válasszák ki: A munka optimalizálása érdekében például a következőket ajánlja: speciális adalékanyagok használata a beton megszilárdulásának felgyorsítására magas páratartalmú körülmények között; az öntést a csapadék legalacsonyabb valószínűségű időszakokra tervezni; ideiglenes menedékhelyeket szervezni a szerkezethez, ami kedvezőtlen időjárási körülmények között is 4-4,5 napra csökkenti a munkaidőt.



Ábra: 1.1-4 Az analitika fő típusai: a múltbeli leírástól az automatizált döntéshozatalig.

A teljes körű digitális átalakulás, amely magában foglalja a rendszeranalitikára és az adatvezérelt irányításra való áttérést, nemcsak kiszervezést, hanem egy hozzáértő belső csapat kialakítását is igényli. Egy ilyen csapat legfontosabb tagjai termékmenedzserek, adatmérnökök, elemzők és fejlesztők, akik szorosan együttműködnek az üzleti egységekkel (4.3-9. ábra). Ez az együttműködés szükséges az intelligens elemző kérdések feltevéséhez és az üzleti döntéshozatali feladatok hatékony paraméterezéséhez. Az információs társadalomban az adatok nem csupán segédeszközöké válnak, hanem az előrejelzés és az optimalizálás alapjává.

Az építőiparban a digitális átalakulás alapjaiban változtatja meg a létesítmények tervezésének, irányításának és üzemeltetésének módját. Ezt a folyamatot az információ digitalizálásaként emlegetik - ahol az építési folyamat minden aspektusa digitalizálódik, és elemzésre alkalmas digitális formába kerül.

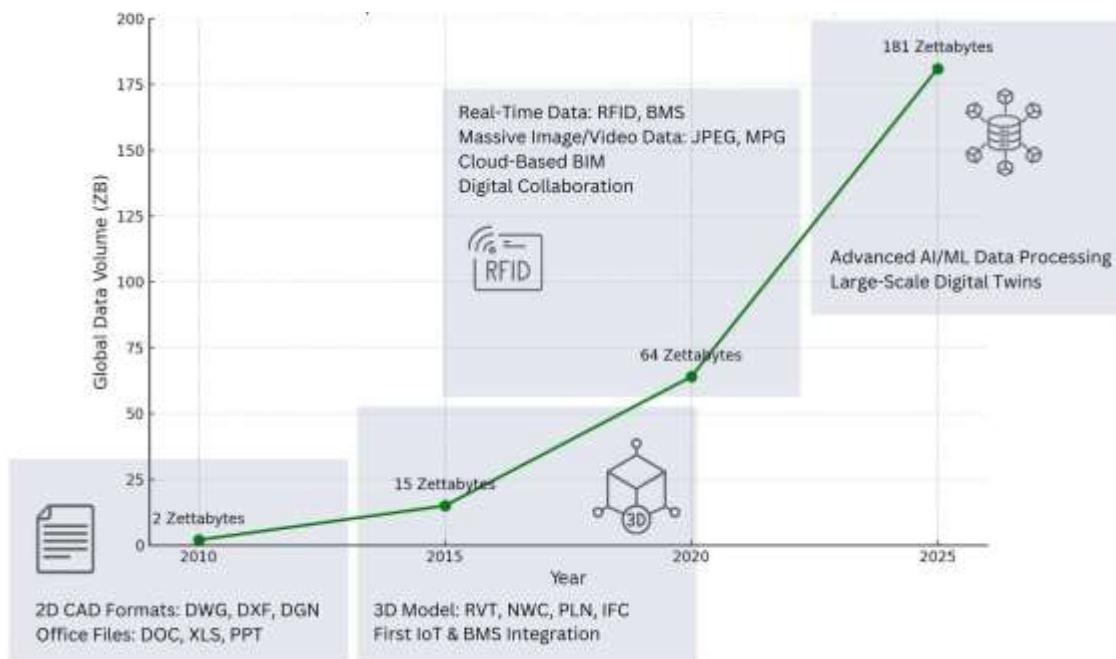
Az építési folyamatokra vonatkozó információk digitalizálása

Az évezredek során az építőiparban rögzített információk mennyisége alig változott, az utóbbi évtizedekben azonban rohamosan nőtt (1.1-5. ábra).

A PwC tanulmánya szerint® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] szerint a világban az összes adat 90%-a az elmúlt két évben jött létre (2015-től). A legtöbb vállalat azonban nem használja ki teljes mértékben ezeket az adatokat, mivel azok vagy sűrűn rendszerekben maradnak, vagy egyszerűen archiválják őket valódi elemzés nélkül.

Az adatmennyiség növekedése az elmúlt években csak felgyorsult, a 2015-ös 15 zettabájtról 2025-re 181

zettabájtra duplázódik [10]. Az építőipari és tervező cégek szerverei nap mint nap megtelnek projektdokumentációkkal, munkatervekkel, számításokkal és kalkulációkkal, pénzügyi jelentésekkel. A 2D/3D -rajzokhoz a DWG, DXF és DGN formátumokat használják, a 3D modellekhez pedig az RVT, NWC, PLN és IFC™ formátumokat. A szöveges dokumentumokat, táblázatokat és prezentációkat DOC, XLSX és PPT formátumban mentjük. Az építkezésről készült videók és képek mellett - MPG és JPEG formátumban - az IoT-elemekből származó valós idejű adatok, az RFID® címkék (azonosítás és nyomon követés) és a BMS épületirányítási rendszerek (felügyelet és ellenőrzés)



ábra 1.1-5 Parabolikus adatnövekedés 2010-2025 ([10] alapján).

Az információ gyors növekedésével az építőiparnak nemcsak az adatok gyűjtésére és tárolására van szüksége, hanem azok ellenőrzésére, érvényesítésére, mérhetőségére és analitikus feldolgozására is. Napjainkban az iparág az információs digitalizáció aktív szakaszán megy keresztül - az építőipari tevékenység minden aspektusának szisztematikus átalakítása elemzésre, értelmezésre és automatizálásra alkalmas digitális formába.

Az információ digitalizálása azt jelenti, hogy az építési projekt és maga az építési folyamat valamennyi egységére és elemére vonatkozó információkat - beleértve azokat is, amelyeket korábban egyáltalán nem tekintettünk információknak - adatformátumba konvertáljuk, hogy az információk számszerűsíthetővé és könnyen elemezhetővé váljanak.

Az építőiparban ez azt jelenti, hogy a projektek minden eleméről és minden folyamatról - a gépek és emberek mozgásától az építkezésen a gépek és emberek mozgásától az építkezésen uralkodó időjárási és éghajlati viszonyokig, az aktuális anyagárakig és a központi banki kamatlábakig - információt kell gyűjteni és digitalizálni, hogy analitikus modelleket lehessen létrehozni.

Ha meg tudod mérni, amiről beszélsz, és számokban is ki tudod fejezni, akkor tudsz valamit a témaáról. De ha nem tudod mennyiségileg kifejezni, akkor a tudásod rendkívül korlátozott és nem kielégítő. Ez lehet egy kiindulópont, de nem a valódi tudományos ismeretek szintje. [11].

- W. Thomson (Lord Kelvin), 1824-1907, brit tudós.

Az információk digitalizálása messze túlmutat az információgyűjtés hagyományos megközelítésén, ahol csak az olyan alapvető mérőszámokat rögzítették, mint a munkaórák vagy a tényleges anyagköltségek. Ma már gyakorlatilag bármilyen esemény átalakítható olyan adatfolyammá, amely fejlett analitikai eszközök és gépi tanulási technikák segítségével mélyreható elemzésre alkalmas. Az építőiparban alapvető váltás történt a papírrajzoktól, Excel-táblázatoktól és szóbeli utasításoktól a digitális rendszerek felé (1.2-4. ábra), amelyekben a projekt minden eleme adatforrássá válik. Még az alkalmazottakat is - a mérnököktől az építkezésen dolgozó munkásokig - ma már digitális változók és adathalmazok gyűjteményének tekintik.

A KPMG "Ismerős kihívások - új megközelítések: Globális építőipari felmérés 2023" című tanulmánya szerint a digitális ikerk, a mesterséges intelligencia (AI) és a Big Data, a projektek jövedelmezőségének javítását elősegítő kulcsfontosságú tényezőként jelennek meg [2].

A modern technológiák nemcsak az információgyűjtést egyszerűsítik, és nagyrészt automatikussá teszik, hanem radikálisan csökkentik az adattárolás költségeit is. Ennek eredményeképpen a vállalatok egyre inkább eltávolodnak a szelektív megközelítéstől, és inkább a teljes információhalmazt tárolják későbbi elemzés céljából (2.1-5. ábra), ami a jövőben potenciális lehetőségeket nyit a folyamatok optimalizálására.

Az információ digitalizálása és a digitalizáció lehetővé teszi az információ rejtett, eddig kiaknázatlan értékének feltárását. Megfelelő szervezés esetén az adatok újrafelhasználhatók, újraérthetők és új szolgáltatásokba és megoldásokba integrálhatók.

A jövőben az információk digitalizálása valószínűleg a dokumentumkezelés teljes automatizálásához, az önmenedzselt építési folyamatok bevezetéséhez és új szakmák - építési adatelemzők, mesterséges intelligencia projektmenedzsment szakértők és digitális mérnökök - megjelenéséhez vezet. Az építési projektek dinamikus információforrásokká válnak, és a döntéshozatal nem intuíciókon vagy szubjektív tapasztalatokon, hanem megbízható és reprodukálható digitális tényeken alapul majd

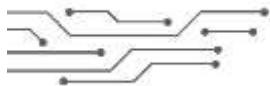
Az információ a 21. század olaja, az analitika pedig a belső égésű motor [12].

- Peter Sondergaard, a Gartner vezető alelnöke®

Az IoT Analytics 2024 [13] szerint az adatkezelésre és -elemzésre fordított globális kiadások a 2023-as 185,5 milliárd dollárról 2030-ra várhatóan drámai mértékben, 16%-os éves növekedési ütemben 513,3 milliárd dollárra nőnek. Azonban nem minden összetevő növekszik ugyanolyan ütemben: az analitika gyorsan növekszik, míg a tárolás növekedése lassul. Az adatkezelési ökoszisztemán belül az analitika biztosítja a leggyorsabb növekedést: az előrejelzések szerint a 2023-as 60,6 milliárd dollárról 2030-ra 227,9 milliárd dollárra nő, ami 27%-os összetett éves növekedési ütemet jelent.

Az információk gyorsuló digitalizációja és az információmennyisége gyors növekedése miatt az építési projektek és a vállalatok vezetésének szisztematikusan tárolnia, elemeznie és feldolgozna kell a különböző, gyakran heterogén adatokat. E kihívásra válaszul az iparág az 1990-es évek közepétől kezdődően nagymértekben áttért a dokumentáció elektronikus létrehozására, tárolására és kezelésére - a táblázatoktól és tervezési rajzokig és szerződésekig.

A hagyományos papíralapú dokumentumokat, amelyek aláírást, fizikai tárolást, rendszeres felülvizsgálatot és szekrényekben való archiválást igényelnek, fokozatosan felváltják a digitális rendszerek, amelyek strukturált módon tárolják az adatokat - speciális alkalmazások adatbázisaiban.



FEJEZET 1.2.

TECHNOLÓGIÁK ÉS IRÁNYÍTÁSI RENDSZEREK A MODERN ÉPÍTŐIPARBAN

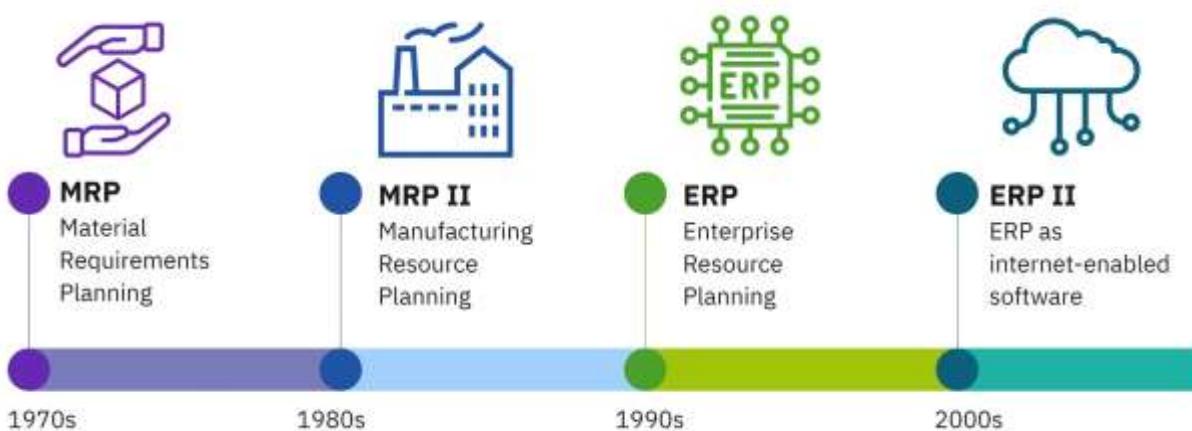
A digitális forradalom és a moduláris MRP/ERP -rendszerek megjelenése.

A modern digitális adattárolás és -feldolgozás korszaka az 1950-es években a mágnesszalagok megjelenésével kezdődött, amely nagy mennyiségű információ tárolásának és felhasználásának lehetőségét nyitotta meg. A következő áttörést a lemezmeghajtók megjelenése jelentette, amely gyökeresen megváltoztatta az építőipar adatkezelési megközelítését.

Az adattárházak fejlődésével számos vállalat lépett be a megoldások piacára, és moduláris szoftverek fejlesztésébe kezdett az adatok létrehozására, tárolására, feldolgozására és a rutinfeladatok automatizálására

Az információk és eszközök exponenciális növekedése miatt olyan integrált, moduláris megoldásokra van szükség, amelyek nem egyedi fájlokkal dolgoznak, hanem segítenek az adatáramlás kezelésében és ellenőrzésében a folyamatok és projektek között.

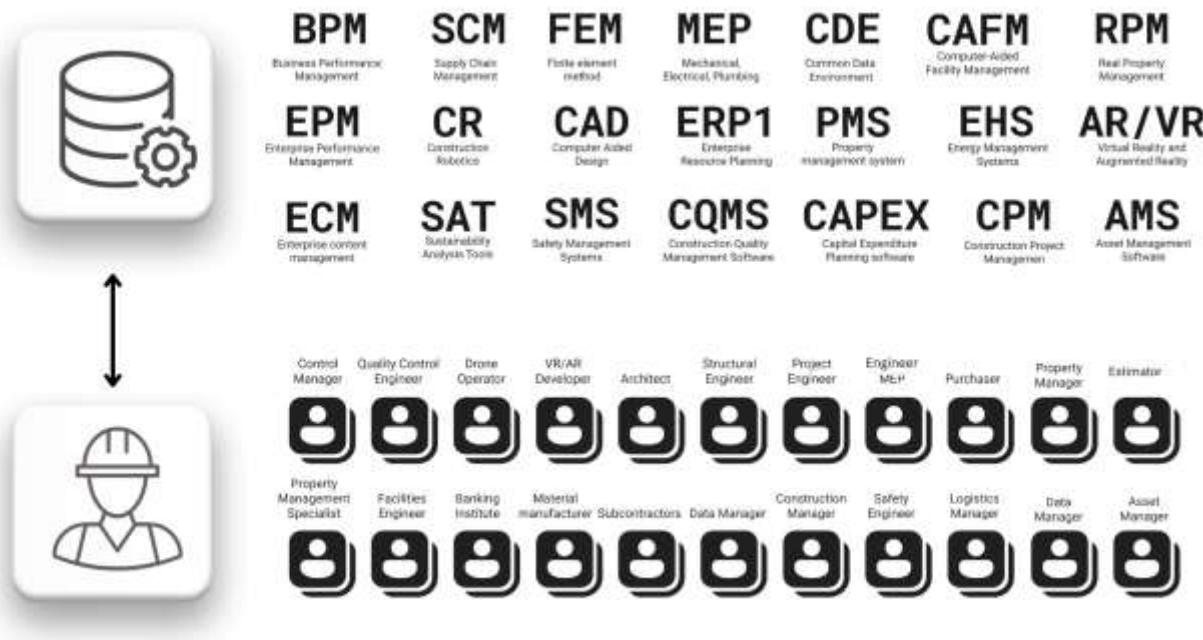
Az első átfogó platformeszközöknek nemcsak dokumentumokat kellett tárolniuk, hanem dokumentálniuk kellett a folyamatokban az összes változtatási kérelmet és műveletet: ki kezdeményezte őket, mi volt a kérelem hatóköré, és végül mi került rögzítésre értékként vagy attribútumként. Ehhez olyan rendszerre volt szükség, amely képes volt a pontos számítások és a meghozott döntések nyomon követésére (1.2-1. ábra). Ilyen platformok voltak az első MRP (Material Requirements Planning) és ERP (Enterprise Resource Planning) rendszerek, amelyek az 1990-es évek elejétől váltak népszerűvé [14]



1.2-1 ábra Az adattárolási technológia fejlődése vezetett az ERP -rendszer megjelenéséhez az 1980-as években.

Az első MRP - és ERP - rendszerek megalapozták a digitalizáció korszakát az üzleti folyamatok és az építési projektek irányításában. Az eredetileg a kulcsfontosságú üzleti folyamatok automatizálására tervezett moduláris rendszereket végül további, rugalmasabb és adaptívabb szoftvermegoldásokkal integrálták.

Ezek a kiegészítő megoldások az adatfeldolgozásra és a projekt-tartalomkezelésre készültek (1.2-2. ábra), vagy a nagy rendszerek egyes moduljait váltották ki, vagy hatékonyan kiegészítették azokat, kiterjesztve a teljes rendszer funkcionalitását.



Ábra: 1.2-2 Az új szoftvermegoldások menedzserei egész sereget vonzották be az üzletbe az adatáramlások kezelésére.

Az elmúlt évtizedekben a vállalatok sokat fektettek a moduláris rendszerekbe [15], mivel ezeket hosszú távú integrált megoldásoknak tekintették.

A Software Path 2022-re vonatkozó jelentése [16] szerint az ERP -rendszer egy felhasználóra jutó átlagos költségvetése 9000 dollár. Átlagosan a vállalat alkalmazottainak mintegy 26%-a használ ilyen rendszereket. Így egy 100 felhasználóval rendelkező szervezet esetében az ERP bevezetésének teljes költsége megközelítőleg 900 000 USD-t tesz ki.

A védett, zárt, moduláris megoldásokba történő beruházások egyre kevésbé indokoltak a modern, rugalmas és nyílt technológiák gyors fejlődésének fényében. Ha már történtek ilyen beruházások, fontos, hogy objektíven újraértekeljük a meglévő rendszerek szerepét: vajon hosszú távon továbbra is nélkülözhetetlenek maradnak-e, vagy funkcióik felülvizsgálhatók és hatékonyabban és átláthatóbban megvalósíthatók.

A mai moduláris adatplatformok egyik fő problémája, hogy az adatkezelést zárt alkalmazásokon belül központosítják. Ennek eredményeképpen az adatok - a vállalat legfontosabb eszközei - a konkrét szoftvermegoldások függvényévé válnak, nem pedig fordítva. Ez korlátozza az információk újrafelhasználását, megnehezíti a migrációt, és csökkenti az üzleti agilitást a gyorsan változó digitális környezetben.

Ha valószínű, hogy a zárt moduláris architektúra értéke vagy jelentősége a jövőben csökkeni fog, akkor érdemes a ma felmerült költségeket elsüllyedt költsékként elismerni, és a nyitottabb, skálázhatóbb és alkalmazkodóbb digitális ökoszisztemá felé történő stratégiai elmozdulásra összpontosítani.

A védett szoftvereket az jellemzi, hogy a fejlesztő kizárolagos ellenőrzést gyakorol a forráskód és az ilyen megoldások használata során létrehozott felhasználói adatok felett. A nyílt forráskódú szoftverekkel ellentétben a felhasználók nem férnek hozzá az alkalmazás belső szerkezetéhez, és nem tudják azt önállóan felülvizsgálni, módosítani vagy saját igényeikhez igazítani. Ehelyett olyan licenceket kell vásárolniuk, amelyek a szoftver használatára jogosítják fel őket a gyártó által meghatározott korlátokon belül.

A modern adatközpontú megközelítés más paradigmát kínál: az adatot fontos stratégiai eszköznek kell tekinteni - független, tartós és a konkrét szoftvermegoldásuktól független. Az alkalmazások viszont puszt a datateszközökkel válnak, amelyek szabadon cserélhetők anélkül, hogy a kritikus információk elvesztésének kockázata fennállna.

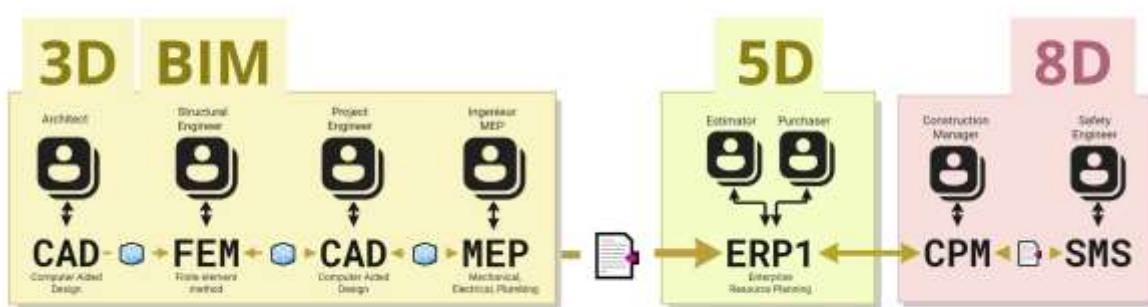
Az ERP és MRP rendszerek kifejlesztése az 1990-es években (1.2-1. ábra) hatékony eszközöket biztosított a vállalkozások számára a folyamatok irányításához, de nem szándékolt következménye volt az is, hogy jelentősen megnövekedett az információáramlás fenntartásában részt vevő személyek száma. Ahelyett, hogy automatizálták és egyszerűsítették volna az operatív feladatokat, ezek a rendszerek gyakran a komplexitás, a bürokrácia és a belső informatikai erőforrásuktól való függőség új szintjeit teremtették meg.

Adatkezelési rendszerek: az adatbányászattól az üzleti kihívásokig

A mai vállalatoknak több adatkezelési rendszer integrálására van szükségük. Az adatkezelési rendszerek kiválasztása, e rendszerek bőlcs kezelése és a különböző adatforrások integrálása kritikus fontosságúvá válik az üzleti teljesítmény szempontjából.

A 2020-as évek közepén több száz (nagy építőipari vállalatknál akár több ezer) különböző rendszerrel találkozhatunk (1.2-3. ábra), amelyeknek összhangban kell működniük ahhoz, hogy az építési folyamat minden aspektusa zökkenőmentesen és egységesen működjön.

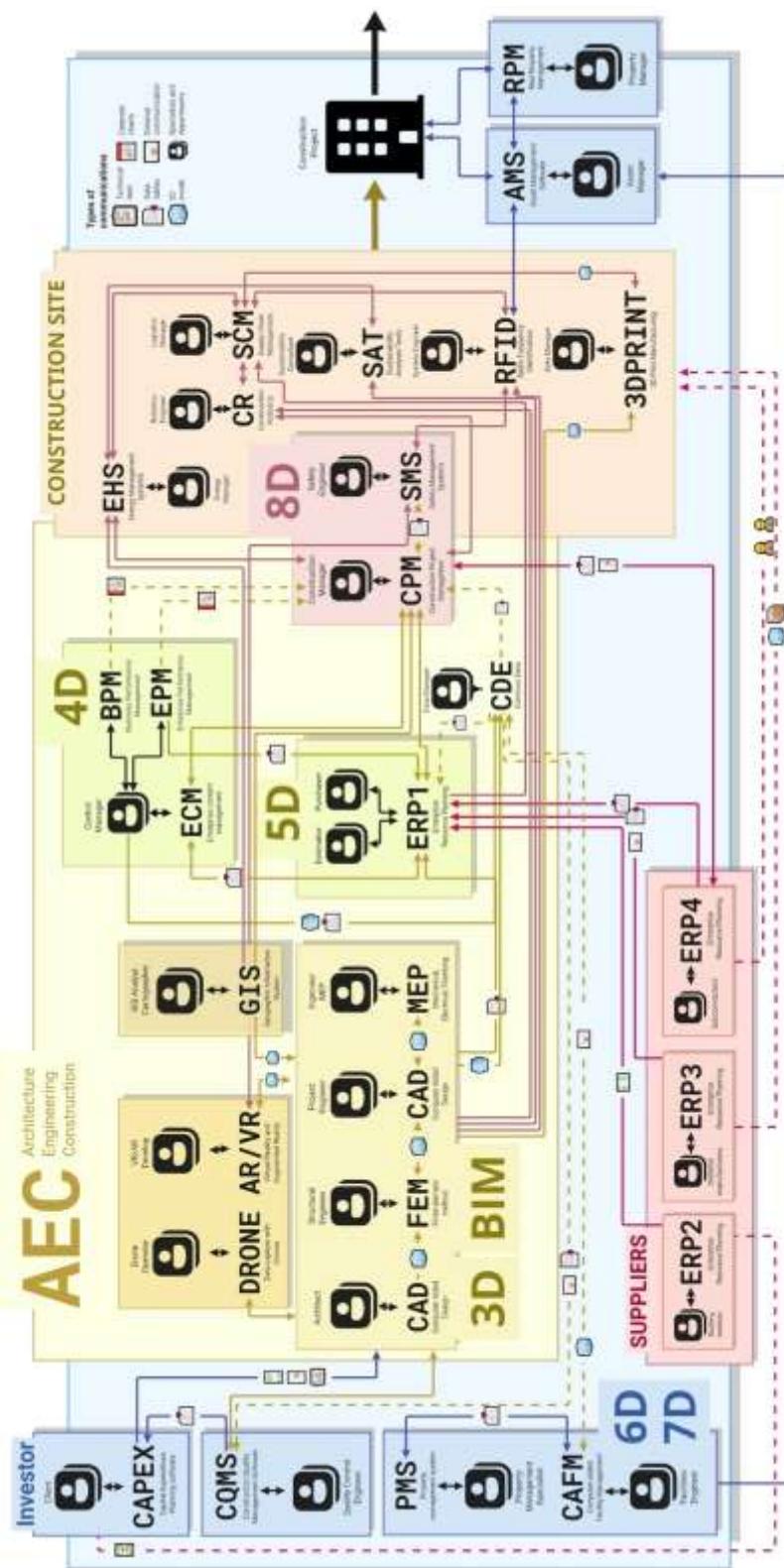
A Deloitte 2016-os tanulmánya szerint -® "Data-Driven Management in Digital Capital Projects" - az átlagos építőipari szakember naponta 3,3 szoftveralkalmazást használ, de ezek közül csak 1,7 van egymással integrálva [17].



Ábra: 1.2-3 minden üzleti rendszernek szüksége van egy professzionális csapatra és egy felelős vezetőre a minőségi adatkezeléshez.

Az alábbiakban felsoroljuk az építőiparban működő közepes és nagyvállalatok számára népszerű rendszereket, amelyeket a hatékony építési projektmenedzsmentben használnak:

- **ERP (Enterprise Resource Planning)** - biztosítja az üzleti folyamatok integrációját, beleértve a könyvelést, a beszerzést és a projektmenedzsmentet.
- **CAPEX (Capital Expenditure Planning Software)** - az építési projektekbe történő pénzügyi beruházások költségvetés-tervezésére és kezelésére szolgál, segít meghatározni a befektetett eszközök és a hosszú távú eszközökbe történő beruházások költségeit.
- **A CAD (Computer-Aided Design) és a BIM (Building Information Modeling)** - a projektek részletes és pontos műszaki rajzainak és 3D-modelljeinek elkészítésére szolgál. E rendszerek középpontjában a geometriai információkkal való munka áll.
- **MEP (Mechanical, Electrical, Plumbing)** - Mérnöki rendszerek, amelyek mechanikai, elektromos és vízvezeték-szerelési komponenseket tartalmaznak, és részletezik a projekt belső "keringési" rendszerét.
- **GIS (földrajzi információs rendszerek)** - terepelemzésre és tervezésre használják, beleértve a kartográfiát és a térbeli elemzést.
- **CQMS (építési minőségirányítási szoftver)** - biztosítja, hogy az építési folyamatok megfeleljenek a megállapított szabványoknak és előírásoknak, és segít a hibák kiküszöbölésében.
- **CPM (építési projektmenedzsment)** - magában foglalja az építési folyamatok tervezését, koordinálását és ellenőrzését.
- **CAFM (Computer-Aided Facility Management)** - épületirányítási és karbantartási rendszerek.
- **Az SCM-re (Supply Chain Management)** a beszállítók és az építkezés közötti anyag- és információáramlás optimalizálása érdekében van szükség.
- **EPM (Enterprise Performance Management)** - célja az üzleti folyamatok és a teljesítmény javítása.
- **AMS (Asset Management Software)** - a berendezések és az infrastruktúra használatának, kezelésének és karbantartásának optimalizálására szolgál az eszközök teljes életciklusa során.
- **RPM (Ingatlankezelés)** - az épületek és földterületek, valamint a kapcsolódó erőforrások és eszközök kezelésével és üzemeltetésével kapcsolatos feladatokat és folyamatokat foglalja magában.



1.2-4. ábra A rendszerek összekapcsolhatósága, amely összekapcsolja a vállalat folyamatait a különböző részlegek közötti információáramlással.

- **CAE (Computer-Aided Engineering)** - Számítógéppel támogatott mérnöki munka, amely magában foglalja a számítási és szimulációs rendszereket, mint például a végeelem-elemanalízis (FEA) és a számítási áramlástan (CFD).
- **CFD (Computational Fluid Dynamics)** - Számítógépes áramlástan, folyadék- és gázáramlások modellezése. CAE alkategória.
- **CAPP (Számítógépes folyamattervezés)** - Számítógépes folyamattervezés. Útvonal- és folyamatábrák létrehozására szolgál.
- **CAM (Computer-Aided Manufacturing)** - számítógéppel támogatott gyártás, CNC-gépek vezérlőprogramjainak generálása.
- **PDM (Product Data Management)** - Termékkadat-kezelés, a műszaki dokumentáció tárolására és kezelésére szolgáló rendszer.
- **A MES (Manufacturing Execution System)** egy valós idejű gyártási folyamatirányító rendszer.
- **PLM (Product Lifecycle Management)** - egy projektelem életciklusának kezelése, integrálja a PDM, CAPP, CAM és más rendszereket a teljes termékellenőrzéshez a fejlesztéstől a selejtezésig.

Ezek és számos más, különböző szoftvermegoldásokat tartalmazó rendszer a modern építőipar szerves részévé vált (1.2-4. ábra). Ezek a rendszerek alapvetően olyan speciális adatbázisok, amelyek intuitív interfésszel rendelkeznek, és a tervezés és az építés minden szakaszában hatékony adatbevitelt, feldolgozást és elemzést biztosítanak. A digitális eszközök egymással való integrációja nemcsak a munkafolyamatok optimalizálásában segít, hanem jelentősen javítja a döntések pontosságát is, ami pozitív hatással van a projekt megvalósításának időzítésére és minőségére.

Az esetek felében azonban nincs integráció. A statisztikák szerint csak minden második alkalmazás vagy rendszer integrálódik más megoldásokkal [17]. Ez a digitális környezet folyamatos széttagoltságát jelzi, és hangsúlyozza, hogy nyílt szabványok és egységes interfések kidolgozására van szükség az építési projekteken belüli végponttól végpontig tartó információcsere biztosítása érdekében.

A modern vállalatok számára az integráció egyik fő kihívása továbbra is a digitális rendszerek nagyfokú összetettsége és a hatékony információkereséshez és értelmezéshez szükséges felhasználói kompetencia követelményei. minden egyes, a vállalatnál bevezetett rendszer támogatására szakemberekből álló csapat alakul, amelynek élén egy kulcsmenedzser áll (1.2-2. ábra).

A kulcsfontosságú rendszermenedzser döntő szerepet játszik az adatáramlás helyes irányításában, és felelős a végső információ minőségéért, ahogy évezredekkel ezelőtt az első menedzserek feleltek a papiruszra vagy agyagtáblára írt számokért.

Ahhoz, hogy a széttagolt információáramlást irányítási eszközzé lehessen alakítani, elengedhetetlen az adatok szisztematikus integrálásának és kezelésének képessége. Ebben az architektúrában a vezetőknek egy egységes hálózat elemeiként kell viselkedniük - mint egy micélium, amely a vállalat egyes részeit összefüggő élő organizzmussá kapcsolja össze, amely képes alkalmazkodni és fejlődni.

Vállalati micélium: hogyan kapcsolódnak az adatok az üzleti folyamatokhoz

Az adatok alkalmazásokba és adatbázisokba történő integrálásának folyamata a különböző forrásokból, többek között különböző osztályokról és szakemberektől származó információk összesítésén alapul (1.2-4. ábra). A szakemberek keresik a releváns adatokat, feldolgozzák azokat, és további felhasználás céljából

átviszik a rendszereikbe és alkalmazásaikba.

Minden egyes vállalati rendszer, amely eszközök, technológiák és adatbázisok összességeből áll, egy tudásfa, amely a történelmi adatok talajában gyökerezik, és új gyümölcsöket terem, hogy kész megoldások: dokumentumok, számítások, táblázatok, grafikonok és műszerfalak formájában új gyümölcsöket teremjen (1.2-5. ábra). A vállalat rendszerei, mint egy erdőben a fák, kölcsönhatásban vannak és kommunikálnak egymással, összetett és jól strukturált rendszert alkotva, amelyet szakértő vezetők támogatnak és irányítanak.

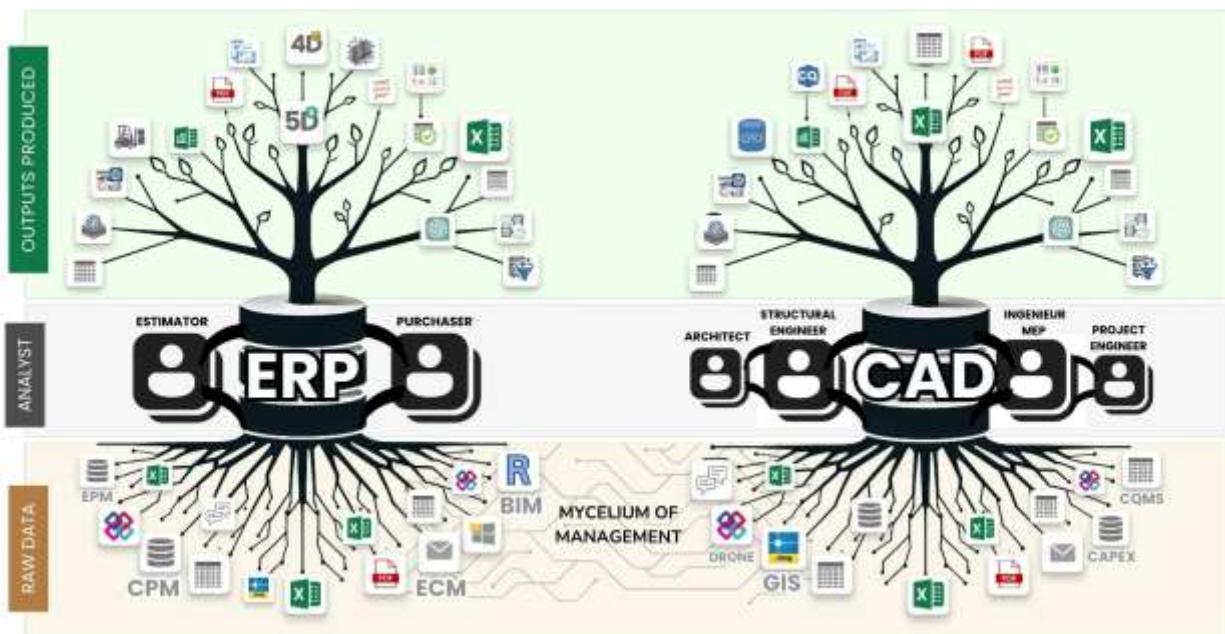
Egy vállalat információkeresési és -átviteli rendszere úgy működik, mint egy komplex erdei ház, amely fákból (rendszerök) és micéliumgombákból (menedzserek) áll, amelyek vezetőként és újrahasznosítóként működnek, biztosítva az információ átadását és áramlását a megfelelő rendszerekhez. Ez segít fenntartani az adatok egészséges és hatékony áramlását és elosztását a vállalaton belül.

A szakértők, akárcsak a gyökerek, a projekt kezdeti szakaszában felszívják a nyers adatokat, és a vállalati ökoszisztemája tápanyagává alakítják azokat. Az adat- és tartalomkezelő rendszerek (1.2-4. ábra - ERP, CPM, BIM, stb.) hatékony információs autópályaként működnek, amelyeken keresztül ez a tudás a vállalat minden szintjén keresztüláramlik.

Ahogy a természetben, ahol az ökoszisztemá minden egyes eleme a maga szerepét tölti be, úgy a vállalati üzleti környezetben is minden egyes folyamat résztvevője - a mérnöktől az elemzőig - hozzájárul az információs környezet növekedéséhez és termékenységéhez. Ezek a rendszerszintű "adatfák" (1.2-5. ábra) nem csupán az információgyűjtés mechanizmusai, hanem a vállalat fenntarthatóságát biztosító versenylőök.

Az erdei ökoszisztemák meglepően pontosan tükrözik a digitális vállalati struktúrák szerveződését. Az erdő többszintű szerkezetéhez hasonlóan - az aljnövényzettől a fák tetejéig - a vállalatirányítás is feladatait rendel a felelősségi szintekhez és a funkcionális részlegekhez.

A fák mély és elágazó gyökerei ellenálló képességet és hozzáférést biztosítanak a tápanyagokhoz. Hasonlóképpen, a szilárd szervezeti struktúra és a minőségi adatokkal való munka stabil folyamatai támogatják a vállalat teljes információs ökoszisztemáját, hozzájárulva a vállalat fenntartható növekedéséhez és fejlődéséhez még a (nagy széllel járó) piaci instabilitás és válságok időszakaiban is.



Ábra 1.2-5 Az adatok különböző rendszereken keresztül történő integrálása olyan, mint egy micélium, amely a vezetőket és a szakembereket egyetlen információs hálózattá kapcsolja össze.

Az üzleti életben a méretarányok modern értelmezése fejlődött. Ma már egy vállalat értékét nemcsak a látható része - a záródokumentumok és jelentések formájában megjelenő "koronák" - határozzák meg, hanem a minőségileg összegyűjtött és szisztematikusan feldolgozott adatok "gyökérrendszerének" mélysége is. Minél több információt lehet összegyűjteni és feldolgozni, annál magasabb lesz az üzleti érték. Azok a vállalatok, amelyek módszeresen felhalmozzák a már feldolgozott adatok "komposztját", és képesek hasznos felismeréseket kinyerni belőle, stratégiai előnyre tesznek szert

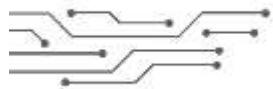
A történelmi információk újfajta tőkévé válnak, amely lehetővé teszi a növekedést, a folyamatok optimalizálását és a versenyelőnyt. Az adatvezérelt világban nem az nyer, akinek több van, hanem az, aki többet tud.

Az építőipar számára ez a valós idejű projektmenedzsmentre való áttérést jelenti, ahol minden folyamat - a tervezéstől a beszerzésen át a kivitelezők koordinálásáig - releváns, naponta frissített adatokon alapul. A különböző forrásokból származó információk integrálása (ERP-rendszerek, CAD-modellek, érzékelők IoT az építkezéseken, RFID) lehetővé teszi a pontosabb előrejelzések készítését, a változásokra való gyors reagálást és a naprakész adatok hiánya miatti késedelmek elkerülését.

A McKinsey & Company Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]) című tanulmánya szerint a jövő sikeres vállalatai a stratégiai döntésektről az operatív interakcióig minden kulcsfontosságú szempontból az adatokra támaszkodnak majd.

Az adatok nem lesznek többé csak elemzési eszközök, hanem az összes üzleti folyamat szerves részévé válnak, átláthatóságot, ellenőrzést és a vezetés automatizálását biztosítva. A alatt található adatvezérelt lépés lehetővé teszi a szervezetek számára, hogy minimalizálják az emberi tényező hatását, csökkentsék a működési kockázatokat, és növeljék a döntéshozatal átláthatóságát és hatékonyságát.

A 21. század a feje tetejére állítja a gazdasági paradigmát: míg korábban az olajat "fekete aranynak" nevezték a gépek és a közlekedés meghajtására való képessége miatt, ma az idő szorításában a történelmi adatok új stratégiai erőforrássá válnak, és nem a gépeket, hanem az üzleti életet irányító döntési algoritmusokat hajtják meg.



FEJEZET 1.3.

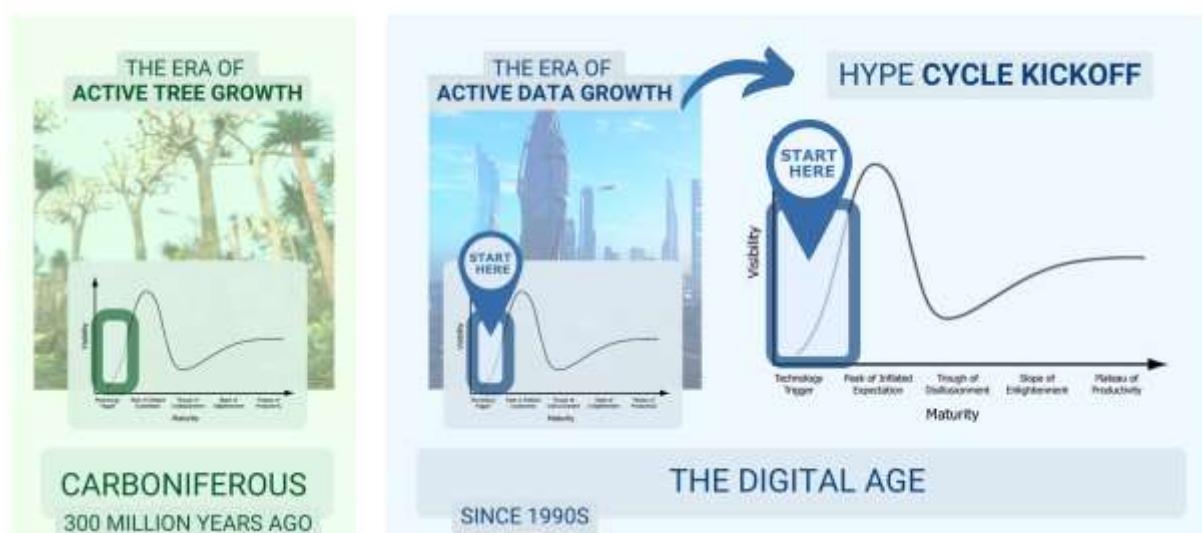
A DIGITÁLIS FORRADALOM ÉS AZ ADATROBBANÁS

Az adatboom kezdete, mint evolúciós hullám

Az építőiparban példátlan információs robbanás tapasztalható. Ha az üzleti életre úgy gondolunk, mint egy tudásfára (1.2-5. ábra), amelyet adatok táplálnak, akkor a digitalizáció jelenlegi szakasza a növényzet gyors növekedéséhez hasonlítható a karbonkorszakban, amikor a bioszférát a biomassza gyors felhalmozódása alakította át (1.3-1. ábra).

A globális digitalizációval az építőiparban az információ mennyisége évente megduplázódik. A modern technológia lehetővé teszi az adatok háttérben történő gyűjtését, valós idejű elemzését és olyan mértékű felhasználását, amely nemrég még lehetetlennek tűnt.

A Gordon Moore (az Intel® társalapítója) által megfogalmazott Moore-törvény szerint az integrált áramkörök sűrűsége és bonyolultsága, valamint a feldolgozott és tárolt adatok mennyisége körülbelül kétévente megduplázódik [19].



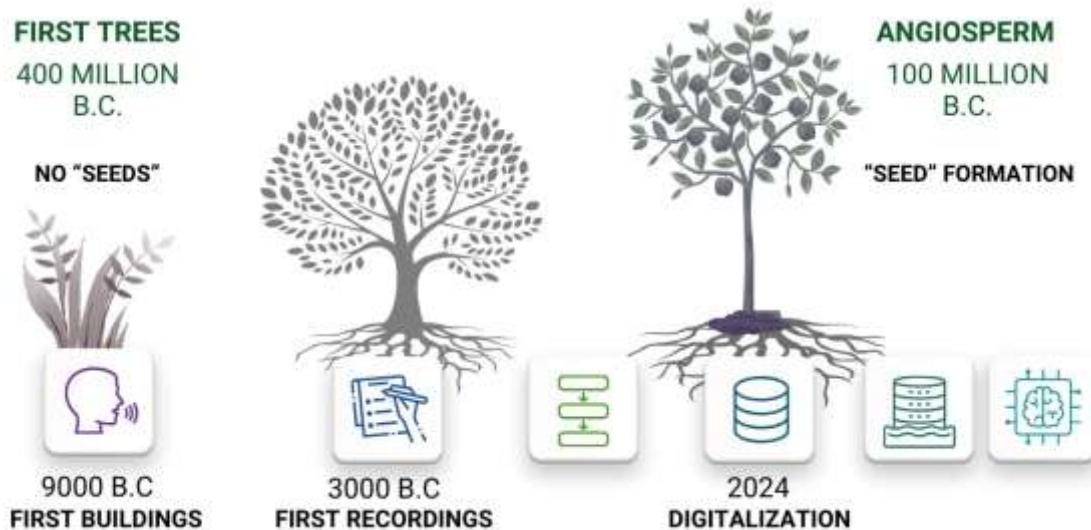
Ábra 1.3-1 A digitalizáció kezdete az adatok exponenciális növekedéséhez vezetett, hasonlóan a szénkorszakban a növényzet robbanásszerű növekedéséhez.

Míg az ősi megalitikus építmények, mint például a törökországi Göbekli Tepe, nem hagytak maguk után újrafelhasználásra alkalmas dokumentált tudást, addig ma a digitális technológiák lehetővé teszik az információk felhalmozását és újrafelhasználását. Ezt a spórás növényekről a magról a magról a magról a magról a magról (angiospermák) való evolúciós átmenethez hasonlíthatjuk: a mag megjelenése adta meg az élet széles körű elterjedését a bolygón. (1.3-2. ábra).

Hasonlóképpen, a múltbeli projektek adatai egyfajta "digitális magokká" válnak - DNS-tudáshordozókká, amelyek skálázhatók és felhasználhatók új projektekben és producerek. A modern mesterséges intelligenciaeszközök - a gépi tanulás és a nagy nyelvi modellek (LLM), mint például a ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok - megjelenése lehetővé teszi az adatok automatikus kinyerését,

értelemezését és új kontextusokban való alkalmazását

Ahogyan a magok forradalmasították az élet elterjedését egy kezdetben élettelen bolygón, az "adatmagok" az új információs struktúrák és tudás automatikus megjelenésének alapjává válnak, lehetővé téve a digitális ökoszisztemák önálló fejlődését és alkalmazkodását a változó felhasználói igényekhez.



1.3-2. ábra A digitális "adatmagok" ugyanolyan evolúciós szerepet játszanak, mint az angiospermák, a Föld ökoszisztemáját átalakító virágos növények.

Az építőipar új korszakának küszöbén állunk, ahol az adatok robbanásszerű növekedése és az "adatmagok" - a múltbeli és folyamatban lévő projektekből származó strukturált információk - aktív terjesztése képezi az ágazat digitális jövőjének alapját. Ezek "beporzása" a nagy adatmennyiségek nyelvi modellekben (LLM) keresztül lehetővé teszi számunkra, hogy ne csak megfigyeljük a digitális változásokat, hanem aktívan részt vegyünk az önállóan tanuló, adaptív ökoszisztemák létrehozásában. Ez nem evolúció - ez egy digitális forradalom, amelyben az adatok egy új valóság fő építőelemévé válnak

Az építőiparban az építési projektek teljes életciklusa során a különböző szakterületekről származó információknak köszönhetően az adatmennyiségek drámaian növekszik. Ez a hatalmas adathalmozódás a Big Data korszaka felé sodorta az építőipart [20].

- Prof. Hang Yang, Építőmérnöki és Építészeti Tanszék, Wuhan University of Technology, Wuhan.,

Az adatok növekedése az információs korban a természetben zajló evolúciós folyamatokra emlékeztet: ahogyan az erdők fejlődése megváltoztatta a bolygó űsi tájképét, úgy a jelenlegi információs robbanás megváltoztatja az egész építőipar tájképét.

Egy modern vállaltnál keletkező adatmennyiség

Az elmúlt két évben a világ összes létező adatának 90%-át létrehozták [21]. 2023-tól kezdve minden egyes ember, beleértve az építőipari szakembereket is, másodpercenként körülbelül 1,7 megabájt adatot generál [22], és a világ összes adatmennyisége 2023-ban eléri a 64 zettabájtot, és az előrejelzések szerint 2025-re meghaladja a 180 zettabájtot, azaz $180 \cdot 10^{15}$ megabájtot [23].

Ennek az információs robbanásnak történelmi előzménye van: a nyomdagépet Johannes Gutenberg találta fel a XV. században. Alig ötven évvel a bevezetése után megduplázódott a könyvek száma Európában: néhány évtized alatt annyi könyvet nyomtattak ki, mint amennyi kézzel készült a megelőző 1200 évben [24]. Napjainkban még ennél is gyorsabb növekedést tapasztalunk: a világ adatmennyisége háromévente megduplázódik.

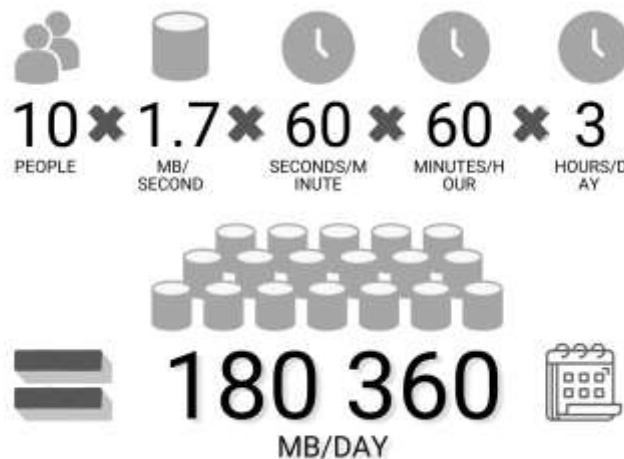
Az adatnövekedés jelenlegi ütemét tekintve az építőipar a következő évtizedekben annyi információt generálhat, mint amennyit egész eddigi történelme során felhalmozott



Ábra: 1.3-3 Az adatok folyamatos növekedéséhez hozzájárul, hogy az egyes alkalmazottak naponta tárolják az adatokat a vállalat szerverein.

A mai építőipari üzleti világban még a kisvállalkozások is hatalmas mennyiségű több formátumú információt generálnak naponta, és még egy kis építőipari vállalat digitális lábnyoma is elérheti a több tíz gigabájtot naponta - a modellektől és rajzoktól kezdve a fényképfelvételeken át a helyszíni érzékelőkig. Ha feltételezzük, hogy minden egyes technikus átlagosan körülbelül 1,7 MB adatot generál másodpercenként, akkor ez körülbelül 146 GB-nak felel meg naponta, vagy 53 TB-nak évente (1.3-3. ábra).

Ha egy 10 fős csapat csak napi 3 órát dolgozik aktívan, a naponta keletkező információ halmozott mennyisége eléri a 180 gigabájtot (1.3-4. ábra).



1.3-4 ábra Egy 10 fős vállalat naponta körülbelül 50-200 gigabájtnyi adatot generál.

Feltételezve, hogy a munkaadatok 30%-a új (a többi felülírózik vagy törlőzik), egy 10 fős cég havonta több száz gigabájtnyi új adatot hozhat létre (a tényleges számok a vállalat tevékenységének típusától függnek)

Világos tehát, hogy nem csak egyre több adatot generálunk, hanem egyre nagyobb szükségünk van az adatok hatékony kezelésére, tárolására és hosszú távú rendelkezésre állására is. És míg korábban az adatok költségmentesen "hevertek" a helyi szervereken, a digitális átalakulással összefüggésben egyre több vállalat kezd felhőmegoldásokat használni információs infrastruktúrájának alapjaként.

Az adattárolás költségei: gazdasági szempontok

Az utóbbi években egyre több vállalat szervezi ki az adattárolást felhőszolgáltatásokba. Ha például egy vállalat adatainak felét a felhőben tárolja, akkor havi 0,015 dolláros átlagos gigabájtonkénti havi ár mellett a tárolási költségei 10-50 dollárral [25] nőhetnek.

Egy kisvállalat számára, tipikus adattermelődési mintákkal, a felhőalapú tárolási költségek több száz és potenciálisan több mint ezer dollár között mozoghatnak havonta (1.3-5. ábra) a néhány év alatt, ami potenciálisan jelentős pénzügyi terhet jelenthet.

A Forrester "Enterprises Outsource Data Storage as Grows Complexity" [26] című tanulmánya szerint, amely 214 technológiai infrastruktúra döntéshozó megkérdezésével készült, a következő adatokat tartalmazza [26], amely 214 technológiai infrastruktúra döntéshozó megkérdezésével készült, a szervezetek több mint egyharmada kiszervezi a tárolást, hogy megbirkózzon az adatműveletek növekvő mennyiségével és összetettségével, és a vállalatok közel kétharmada az előfizetéses modellt részesít előnyben.



Ábra 1.3-5 Az adatok felhőbe költözöttése akár 2000 dollárral is növelheti a havi tárolási költségeket, még egy olyan vállalat esetében is, amelynek mindössze 10 alkalmazottja van.

A helyzetet tovább bonyolítja a felhőalapú technológiák, például a CAD (BIM), CAFM, PMIS és ERP - rendszerek felgyorsult elfogadása, amelyek tovább növelik az adattárolási és feldolgozási költségeket. Ennek következtében a vállalatok kénytelenek keresni a költségoptimalizálás és a felhőszolgáltatóktól való függőségük csökkentésének módjait.

2023 óta, a nagy nyelvi modellek (LLM) aktív fejlesztésével az adattárolás megközelítései változni kezdtek. Egyre több vállalat gondolkodik azon, hogy visszaveszi az adatai feletti ellenőrzést, mivel biztonságosabbá és jövedelmezőbbé válik az információk saját szervereken történő feldolgozása.

Ebben az összefüggésben előtérbe kerül a felhőalapú tárolástól és a csak a szükséges adatok feldolgozásától a vállalati LLM és AI -megoldások helyi telepítése javára történő elmozdulás. Ahogy a Microsoft vezérigazgatója az egyik interjújában [27] rámutatott, ahelyett, hogy több különálló alkalmazásra vagy felhőalapú SaaS megoldásra támaszkodnának a különböző feladatok elvégzéséhez, az AI-ügynökök adatbázisokban fogják kezelni a folyamatokat, automatizálva a különböző rendszerek funkcióját.

[...] a régi megközelítés erre a [adatfeldolgozás] kérdésre a következő volt: ha visszagondolunk arra, hogy a különböző üzleti alkalmazások hogyan kezelték az integrációt, akkor konnektorokat használtak. A vállalatok licenceket adtak el ezekhez a csatlakozókhöz, és az üzleti modell e köré épült. Az SAP [ERP] az egyik klasszikus példa: csak akkor lehetett hozzáérni az SAP-adatokhoz, ha rendelkezett a megfelelő csatlakozóval. Úgy tűnik tehát, hogy valami hasonló fog kialakulni az [AI] ügynöki interakció esetében is [...]. A megközelítés, legalábbis amit mi képviselünk, a következő: úgy gondolom, hogy az üzleti alkalmazások létezésének fogalma valószínűleg össze fog omlani az [AI] ügynökök korában. Mert ha belegondolunk, ezek lényegében adatbázisok egy csomó üzleti logikával

- Satya Nadella, a Microsoft vezérigazgatója, interjú a BG2 csatornának, 2024. [28]

Ebben a paradigmában az adatvezérelt LLM megközelítés túlmutat a klasszikus rendszereken. A mesterséges intelligencia közvetítővé válik a felhasználó és az adatok között (2.2-3. ábra, 2.2-4. ábra),

kiküszöböltre a többszörös közvetítő interfések szükségességét és növelte az üzleti folyamatok hatékonyságát. Az adatokkal való munkának erről a megközelítéséről bővebben a "Káoszból rendet csinálni és a komplexitást csökkenteni" című fejezetben lesz szó.

Miközben a jövő építészete még csak formálódik, a vállalatok már most szembesülnek a múltbeli döntések következményeivel. Az elmúlt évtizedek masszív digitalizációja, amelyet a különböző rendszerek bevezetése és az adatok ellenőrizetlen felhalmozódása kísért, új problémához vezetett - az információ túlterheltséghez.

Az adathalmozás határai: a tömegtől a jelentésig

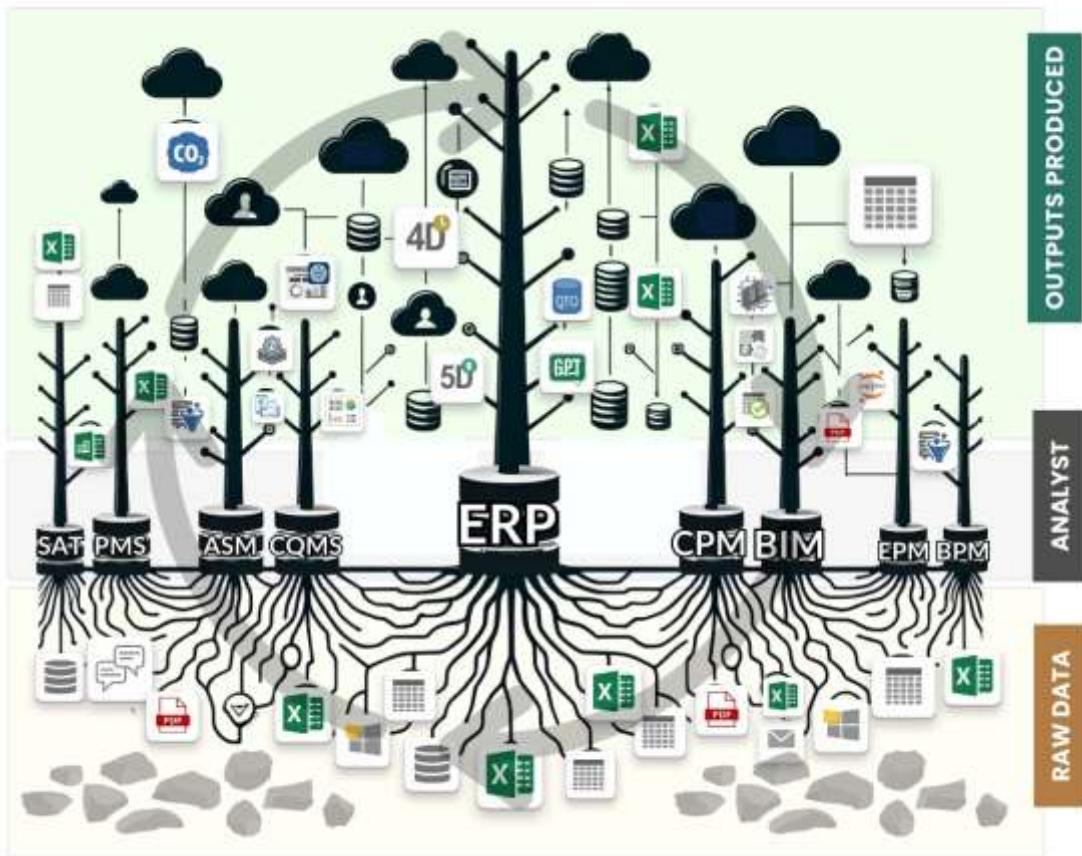
A modern vállalati rendszerek akkor fejlődnek és működnek sikeresen, ha az adatok mennyisége és az alkalmazások száma egyensúlyban van az informatikai részlegek és a vezetők képességeivel. Az elmúlt évtizedekben azonban a digitalizáció az adatok mennyiségének és összetettségének ellenőrizhetetlen növekedéséhez vezetett, ami a vállalatok információs ökoszisztemájában túltelítődési hatást okozott.

Napjainkban a szerverek és a tároló létesítmények soha nem látott mennyiségű feldolgozatlan és többszörösen formázott információnak vannak kitéve, amelyeknek nincs idejük komposztálalakulni, és gyorsan irrelevánsá válnak. A vállalat korlátozott erőforrásai nem tudnak megbirkózni ezzel az áradattal, és az adatok elszigetelt silókban (úgynevezett "silókban") halmozódnak fel, amelyekből a hasznos információk kinyeréséhez kézi feldolgozásra van szükség.

Ennek eredményeképpen, mint egy borostyánnal benőtt és penéssel borított erdő, a modern vállalatirányítási rendszerek gyakran szenvednek az információ túlterheltségétől. Ahelyett, hogy a vállalati ökoszisztemáma szívében tápláló információs humusz lenne, különböző formátumú adatok elszigetelt területei alakulnak ki, ami elkerülhetetlenül az üzleti folyamatok általános hatékonyságának csökkenéséhez vezet.

Az elmúlt 40 év exponenciális adatnövekedésének hosszú időszakát elkerülhetetlenül telítődési és azt követő lehűlési szakasz követi. Amikor a tárolás eléri a határait, minőségi változás következik be: az adatok többé már nem csak tárolási objektumok, hanem stratégiai erőforrások lesznek.

A mesterséges intelligencia és a gépi tanulás fejlődésével a vállalatoknak lehetőségek van az információfeldolgozási költségek csökkentésére, és a mennyiségi növekedésről az adatok minőségi felhasználására való áttérésre. A következő évtizedben az építőiparnak az egyre több adat létrehozásától az adatok szerkezetének, integritásának és elemzési értékének biztosítására kell majd összpontosítania.



1.3-6. ábra Az elszigetelt adatforrások megakadályozzák az adatrendszerök közötti információcserét.

A fő érték már nem az információ mennyiségében rejlik, hanem abban a képességben, hogy automatikusan értelmezzük és a vezetői döntések meghozatalához hasznos, alkalmazott tudássá alakítsuk. Ahhoz, hogy az adatok valóban hasznosak legyenek, megfelelően kell kezelni őket: gyűjteni, ellenőrizni, strukturálni, tárolni és elemezni kell őket a konkrét üzleti feladatok összefüggésében.

A adatelemzési folyamat egy vállaltnál hasonló az erdei fák életciklusához és pusztulásához, valamint az új, fiatal és erős fák megjelenéséhez: az érett fák elpusztulnak, lebomlanak, és az új növekedés táptalajává válnak. A befejezett és befejezetlen folyamatok, amikor befejeződnek, a vállalat információs ökoszisztémájának részévé válnak, és végül az információs humusszá válnak, amely táplálja az új rendszerek és adatok jövőbeli növekedését.

A gyakorlatban azonban ez a ciklus gyakran megszakad. Az organikus megújulás helyett a geológiai rétegekhez hasonló rétegkáosz alakul ki, ahol új rendszerek rétegződnek a régiokre, mély integráció és strukturálódás nélkül. Ennek eredményeképpen széttagolt információs "silók" alakulnak ki, amelyek akadályozzák a tudás áramlását és bonyolítják az adatkezelést.

Következő lépések: az adatelmélettől a gyakorlati változásig

Az építőipari adatok fejlődése az agyagtábláktól a modern moduláris platformokig tartó út. A kihívás már nem az információgyűjtés, hanem egy olyan keretrendszer létrehozása, amely a széttagolt és sokféle adatot stratégiai erőforrássá alakítja. Akár vállalatvezetői, akár mérnöki szerepet tölt be, az adatok értékének és az azokkal való munkavégzésnek a megértése kulcsfontosságú készség lesz a jövőben.

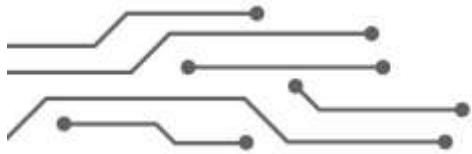
Összefoglalva ezt a részt, érdemes kiemelni a főbb gyakorlati lépéseket, amelyek segítenek a tárgyalt megközelítések alkalmazásában a napi feladatok során:

- Az információáramlás személyes auditjának elvégzése
 - Készítsen egy listát az összes olyan rendszerről és alkalmazásról, amellyel naponta dolgozik.
 - Jelölje meg, hol tölti a legtöbb időt az adatok keresésével vagy kétszeres ellenőrzésével.
 - A legfontosabb információforrások azonosítása
 - Elemezze a jelenlegi alkalmazási környezetet a redundancia és a funkciók megkettőzése szempontjából.
- Törekedjen arra, hogy az analitikai érettségi szintek szerint haladjon a folyamatokon keresztül.
 - Kezdje a feladatokat leíró elemzésekkel (mi történt?)
 - Fokozatosan vezessen be egy diagnosztikát (miért történt ez?)
 - Gondolja át, hogy a folyamatokban hogyan lehet áttérni a prediktív (mi fog történni?) és a preskriptív (mit kell tenni?) elemzésekre.
- Kezdje el strukturálni a munkaadatait
 - Egységes rendszer bevezetése a munkája során gyakran használt fájlok és mappák elnevezésére
 - Sablonok létrehozása gyakran használt dokumentumokhoz és jelentésekhez
 - Rendszeresen archiválja a befejezett projekteket világos struktúrával

Még ha nem is tudja megváltoztatni a csapat vagy a vállalat teljes információs infrastruktúráját, kezdje a saját folyamataival és a minden nap munka apró fejlesztéseiivel. Ne feledje, hogy az adatok valódi értéke nem a mennyiségükben rejlik, hanem abban, hogy képesek vagyunk-e hasznosítható felismeréseket nyerni belőlük. Még a kicsi, de megfelelően strukturált és elemzett adathalmazok is jelentős hatással lehetnek, ha beépülnek a döntéshozatali folyamatokba.

A könyv következő részeiben áttérünk az adatokkal való munka konkrét módszereire és eszközeire, megvizsgáljuk, hogyan lehet a strukturálatlan információkat strukturált halmazokká alakítani, megvizsgáljuk az analitikai automatizálási technológiákat, és részletezzük, hogyan lehet hatékony analitikai ökoszisztemát kiépíteni egy építőipari vállalatnál.





II RÉSZ

HOGYAN FULLAD BELE AZ ÉPÍTŐIPAR AZ ADATKÁOSZBA

A második rész az építőipari vállalatok előtt álló kihívások kritikus elemzésével foglalkozik, amelyekkel a növekvő adatmennyiségek kezelése során szembesülnek. Részletesen megvizsgáljuk az információk széttöredezettségének következményeit és a "silókban tárolt adatok" jelenségét, amely akadályozza a hatékony döntéshozatalt. A HiPPO -megközelítés (Highest Paid Person's Opinion) problémáit és annak az építési projektekkel kapcsolatos vezetői döntések minőségére gyakorolt hatását vizsgálják. Értékelik a dinamikus üzleti folyamatok és növekvő komplexitásuk hatását az információáramlásra és a működési hatékonyságra. Konkrét példák mutatják be, hogy a rendszerek túlzott összetettsége hogyan növeli a költségeket és csökkenti a szervezetek rugalmasságát. Különös figyelmet fordítanak a védett formátumok okozta korlátozásokra és a nyílt szabványok építőipari alkalmazásának kilátásaira. Bemutatásra kerül a mesterséges intelligencián és LLM-alapú szoftver ökoszisztémák felé való elmozdulás koncepciója, amelyek minimalizálják a túlzott összetettséget és a technikai akadályokat.

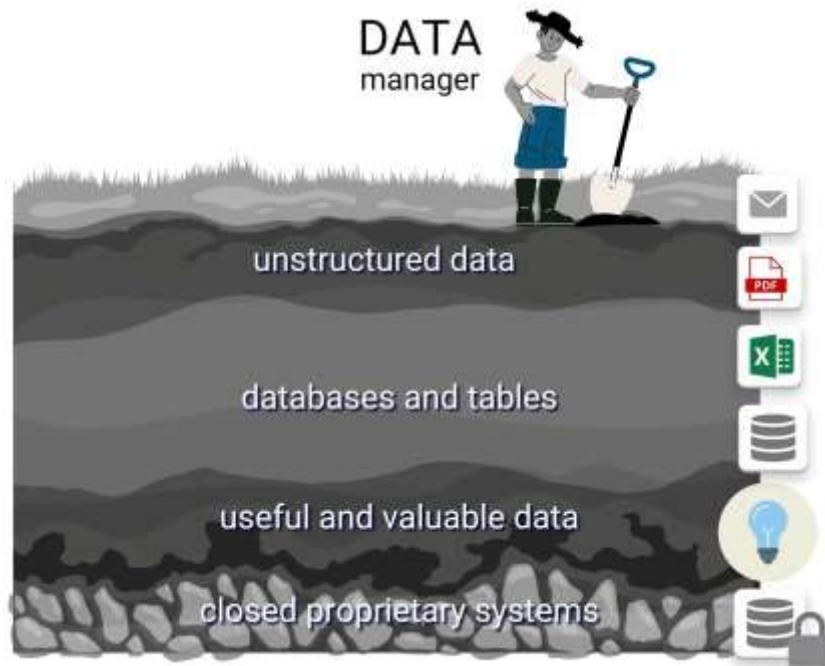
FEJEZET 2.1.

AZ ADATOK SZÉTTÖREDEZETTSÉGE ÉS SILÓK

Minél több eszköz, annál hatékonyabb a vállalkozás?

Első pillantásra úgy tűnhet, hogy a több digitális eszköz nagyobb hatékonyságot eredményez. A gyakorlatban azonban nem ez a helyzet. minden egyes új megoldással, legyen az egy felhőszolgáltatás, egy örökölt rendszer vagy egy újabb Excel-jelentés, a vállalat egy újabb réteggel bővíti digitális tájképét - egy olyan réteggel, amely gyakran nincs integrálva a többivel (2.1-1. ábra).

Az adatok a szénhez vagy az olajhoz hasonlíthatók: évekig tart, amíg felhalmozódnak, a káosz, a hibák, a strukturálatlan folyamatok és az elfelejtett formátumok rétegei alatt tömörülnek. Ahhoz, hogy valóban hasznos információkat nyerjenek ki belőle, a vállalatoknak szó szerint át kell gázolniuk az elavult megoldások és a digitális zaj rétegein.



Ábra 2.1-1 A különböző adatok tagolt rétegeket alkotnak - még az "arany" felismerések is elvesznek a rendszerkomplexitás geológiai szikláiban.

Minden új alkalmazás nyomot hagy maga után: egy fájlt, egy táblázatot vagy egy egész elszigetelt "silót" a szerveren. Az egyik réteg agyag (elavult és elfelejtett adatok), a másik homok (széttagolt táblázatok és jelentések), a harmadik pedig gránit (zárt, védett, nem integrálható formátumok). Idővel a vállalat digitális környezete egyre inkább az információk ellenőrizetlen felhalmozódásának tározójához hasonlít, ahol az érték a vállalat szervereinek mélyén veszik el.

Minden egyes új projekttel és új rendszerrel nemcsak az infrastruktúra válik egyre összetettebbé, hanem a hasznos, minőségi adatokhoz vezető út is. Az értékes "sziklához" való eljutás mélyreható tisztítást, az

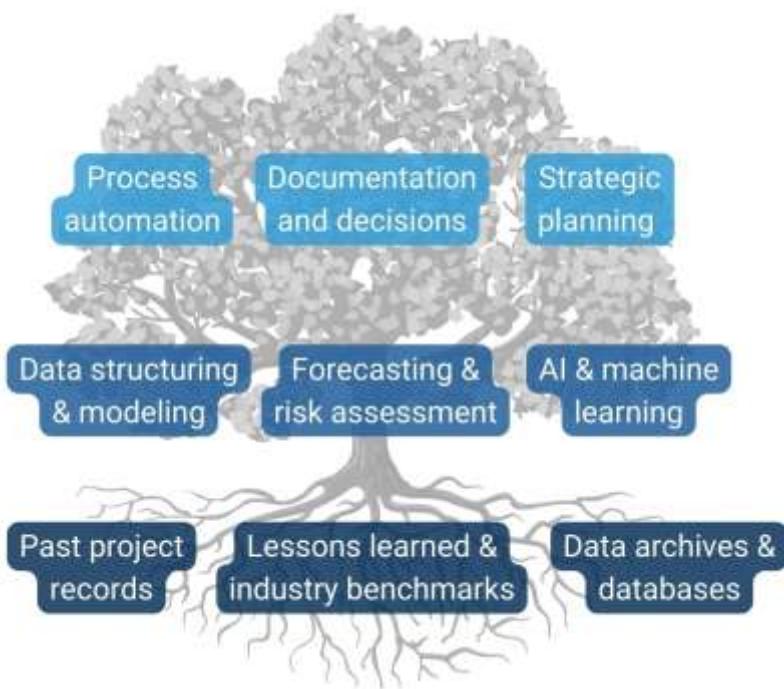
információk strukturálását, "darabolásukat", értelmes darabokba való csoporthozását, valamint a stratégiaileg fontos meglátások analitika és adatmodellezés révén történő kinyerését igényli.

Az adat értékes doleg, és tovább tart, mint maguk a rendszerek [amelyek az adatokat feldolgozzák] [29].

- Tim Berners-Lee, a világháló atyja és az első weboldal megalkotója

Mielőtt az adatok "értékes doleggá" és a döntéshozatal megbízható alapjává válhatnak, alapos előkészítésen kell átesniük. A megfelelő előfeldolgozás az, ami a szétszórta adatokból strukturált élményt, hasznos információs humuszt csinál, amely aztán előrejelzési és optimalizálási eszközzé válik.

Tévhit, hogy az elemzés megkezdéséhez tökéletesen tiszta adatokra van szükség, de a gyakorlatban a piszkos adatokkal való munkavégzés a folyamat alapvető része.



2.1-2. ábra Az adatok az üzleti tevékenység gyökérrendszeré és alapja, amely viszont a döntéshozatali folyamatokon alapul.

A technológia folyamatos fejlődésével az Ön vállalkozásának is előre kell lépnie, és meg kell tanulnia, hogyan teremtsen értéket az adatokból. Ahogyan az olaj- és szénipari vállalatok kiépítik az ásványkincsek kitermeléséhez szükséges infrastruktúrát, úgy a vállalkozásoknak is meg kell tanulniuk, hogyan kezeljék az új információk áramlását saját szervereiken, és hogyan nyerjenek értékes meglátásokat a fel nem használt, nem formázott és elavult adatokból, stratégiai erőforrássá alakítva azokat.

A mezők létrehozása (adattárházak) az első lépés. Még a legerősebb eszközök sem oldják meg az adatelkülönítés és a több formátumú adatok problémáját, ha a vállalatok továbbra is silázott rendszerekben működnek. Ha az adatok egymástól elkülönítve, egymást keresztező és megosztó információk nélkül léteznek, a vállalkozásoknak "adatsiló" hatással kell szembenézniük. Az egységes, konzisztens infrastruktúra helyett a vállalatok kénytelenek erőforrásokat fordítani az adatok összevonására és szinkronizálására.

Adatsilók és azok hatása a vállalati teljesítményre

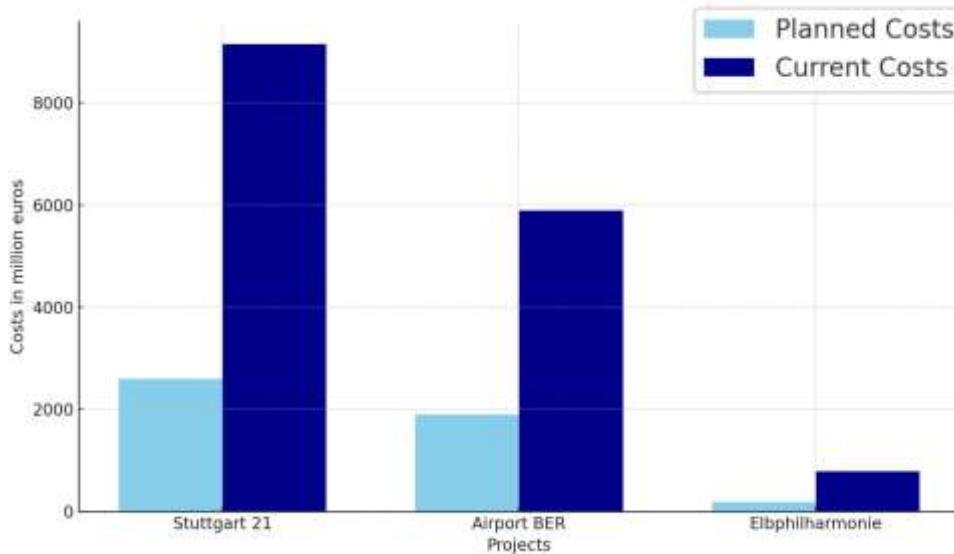
Képzelje el, hogy egy lakótelepet épít, de minden csapatnak saját projektje van. Egyesek falakat építenek, mások kommunikációs csatornákat fektetnek le, mégint mások utakat fektetnek le anélkül, hogy egyeztetnének egymással. Ennek eredményeképpen a csövek nem illeszkednek a falakon lévő nyílásokhoz, a liftaknák nem felelnek meg az emeleteknek, az utakat pedig le kell bontani és újra kell fektetni.

Ez a helyzet nem csak egy hipotetikus forgatókönyv, hanem számos modern építési projekt valósága. A különböző rendszerekkel dolgozó generálisok és alvállalkozók nagy száma miatt, valamint az egységes koordinációs központ hiánya miatt a folyamat végétlen számú jóváhagyás, utómunka és konfliktus sorozatává válik. Mindez jelentős késésekhez és a projekt többszörös költségeihez vezet.

Egy építkezésen a klasszikus helyzet egyszerű: a zsaluzat készen van, de a betonacélok szállítása nem érkezett meg időben. A különböző rendszerekben az információk ellenőrzése során a kommunikáció nagyjából a következőképpen zajlik:

- ⌚ A 20-i építkezésen a **művezető azt** írja a projektvezetőnek: "*Befejeztük a zsaluzást, hol van a betonacél?*".
- ⌚ **Projektmenedzser (PMIS)** a beszerzési osztályhoz: - "*A zsaluzat készen áll. A rendszeremben [PMIS] azt írja, hogy a betonacélnak 18-án kellett volna megérkeznie. Hol vannak a betonacélok?*"
- ⌚ **Ellátási lánc szakértő (ERP)**: - "*Az ERP szerint a szállítás 25-én lesz*".
- ⌚ **Adatmérnök** vagy informatikai osztály (integrációkért felelős): - A PMIS-ben a dátum 18-a, az ERP-ben 25-e. Az ERP és a PMIS között nincs OrderID kapcsolat, így az adatok nincsenek szinkronizálva. Ez egy tipikus példája az információs hiányosságnak.
- ⌚ **A projektmenedzser** a főigazgatónak - "*A szerevények szállítása késik, a helyszín áll, és nem világos, ki a felelős*".

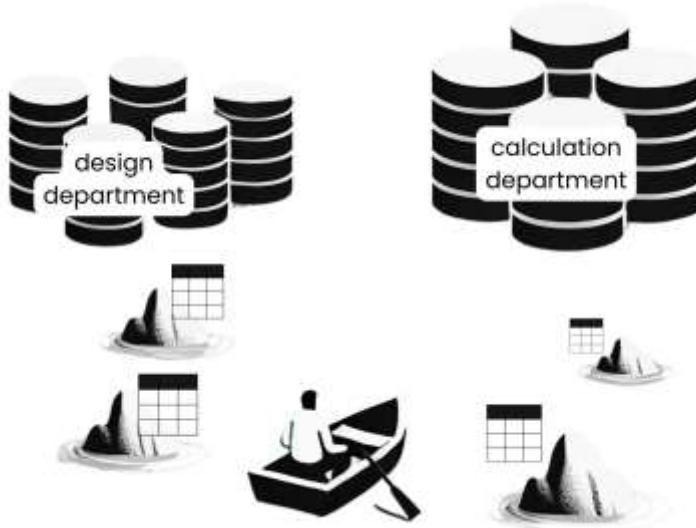
Az incidens oka az volt, hogy az adatok elkülönültek egymástól a különböző rendszerekben. Az adatforrások integrálásával és egységesítésével, az információk egységes tárolásának létrehozásával és az ETL-eszközök (Apache NiFi, Airflow vagy n8n) segítségével történő automatizálásával a rendszerek közötti silók megszüntethetők. Ezeket és más módszereket és eszközöket a könyv későbbi részeiben részletesen tárgyaljuk.



2.1-3 A németországi nagy infrastrukturális projektek tervezett és tényleges költségeinek összehasonlítása.

Ugyanez történik a vállalati rendszerekkel is: először elszigetelt megoldásokat hoznak létre, majd hatalmas költségvetést kell elkölni az integrálásukra és összehangolásukra. Ha az adat- és kommunikációs modelleket már a kezdetektől fogva átgondolták volna, egyáltalán nem lenne szükség integrációra. A silázott adatok káoszt teremtenek a digitális világban, akárcsak egy koordinálatlan építési folyamat.

A KPMG "Cue construction 4.0: Time to make or break" című, 2023-as tanulmánya szerint a vállalatok minden össze 36%-a osztja meg hatékonyan az adatokat az osztályok között, míg 61%-uk komoly problémákkal küzd az elszigetelt adat-silók miatt [30].



Ábra: 2.1-4 Éveken át nehezen gyűjthető adatok halmozódnak fel elszigetelt tároló "silókban", és fennáll a veszélye annak, hogy soha nem használják fel őket.

A vállalati adatokat elszigetelt rendszerekben tárolják, mint a tájban szétszórt egyes fákat. Mindegyik

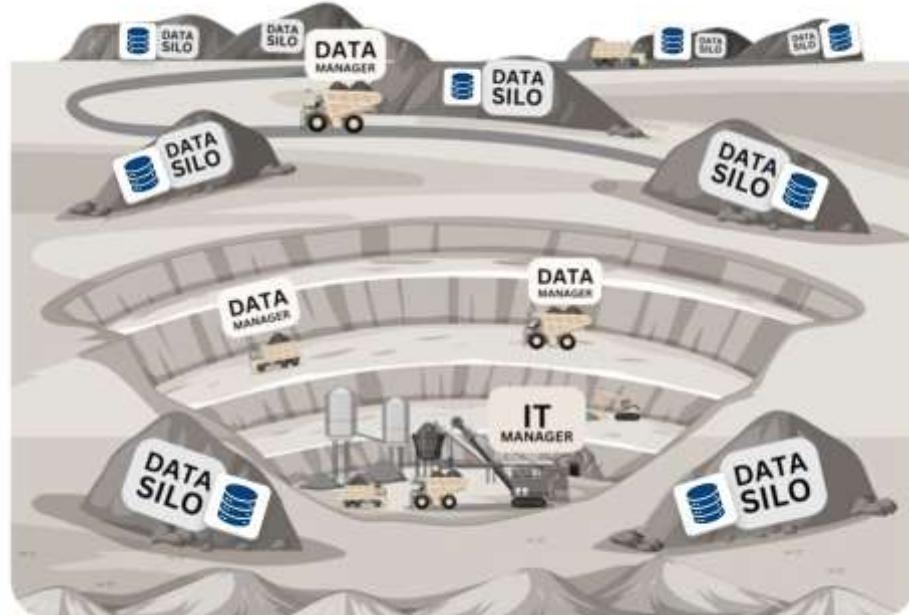
értékes információkat tartalmaz, de a köztük lévő kapcsolatok hiánya megakadályozza egy egységes, összekapcsolt ökosziszterna létrehozását. Ez az elszigeteltség akadályozza az adatáramlást, és korlátozza a szervezet teljes képet alkotó képességét. Ezeknek a az összekapcsolása rendkívül hosszú és összetett folyamat, amelynek során a vezetés szintjén gombamizériát kell növeszteni ahhoz, hogy megtanuljuk, hogyan lehet az egyes információdarabokat a rendszerek között átadni.

A WEF 2016-os tanulmánya szerint a digitális átalakulás egyik fő akadálya a közös adatszabványok hiánya és a széttagoltság.

Az építőipar a világ egyik legelaprózottabb ágazata, és az értéklánc valamennyi résztvevőjének zökkenőmentes együttműködésétől függ [5].

- Világgazdasági Fórum 2016: Az jövőjének alakítása

A tervezők, menedzserek, koordinátorok és fejlesztők gyakran inkább önállóan dolgoznak, elkerülve a koordináció bonyolultságát. Ez a természetes hajlam információs "silók" kialakulásához vezet, amelyekben az adatok különálló rendszerekben vannak elszigetelve. Minél több ilyen elszigetelt rendszer van, annál nehezebb rávenni őket az együttműködésre. Idővel minden egyes rendszer saját adatbázist és egy speciális támogató vezetői részleget kap (1.2-4. ábra), ami tovább bonyolítja az integrációt.



Ábra 2.1-5 Minden rendszer hajlamos arra, hogy létrehozza a saját egyedi silóját adatokból, amelyeket megfelelő eszközökkel kell feldolgozni [31].

A vállalati rendszerek ördögi köre a következőképpen néz ki: a vállalatok komplex ISO leírásokba fektetnek be, majd magas költségekkel szembesülnek az integrációjuk miatt, a fejlesztők pedig, felismervé a rendszerek kombinálásának összetettségét, inkább a zárt ökoszisztemáikban dolgoznak. Mindez növeli az informatikai táj széttöredezettségét, és megnehezíti az új megoldásokra való áttérést (2.1-5. ábra). A vezetők végül az adatsilókat kritizálják, de ritkán elemzik az okokat és a megelőzésük módját. A vezetők

panaszoknak az elavult informatikai rendszerek miatt, de azok cseréje jelentős beruházást igényel, és ritkán hozza meg a várt eredményeket. Ennek eredményeképpen a probléma leküzdésére tett kísérletek is gyakran csak rontanak a helyzeten.

A kapcsolat megszakadásának fő oka, hogy az alkalmazások elsőbbséget élveznek az adatokkal szemben. A vállalatok először különálló rendszereket fejlesztenek ki, vagy kész megoldásokat vásárolnak a szállítóktól, majd ezeket duplikált és inkompatibilis tárolók és adatbázisok létrehozásával próbálják egységesíteni.

A széttöredezetség problémájának megoldása radikálisan új megközelítést igényel - az adatok elsőbbséget élveznek az alkalmazásokkal szemben. A vállalatoknak először adatkezelési stratégiákat és adatmodelleket kell kidolgozniuk, majd olyan rendszereket kell létrehozniuk vagy olyan megoldásokat kell megvásárolniuk, amelyek egységes információkészlettel dolgoznak, ahelyett, hogy új akadályokat állítanának fel.

Olyan új világba lépünk, ahol az adatok fontosabbak lehetnek, mint a szoftverek.

- Tim O'Reilly, az O'Reilly Media, Inc. vezérigazgatója.

A McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) című tanulmánya azt mutatja, hogy az építőipar lemaradásban van más ágazatokhoz képest a digitális átalakulás terén [32]. A jelentés szerint az automatizált adatkezelés és a digitális platformok bevezetése jelentősen javíthatja a termelékenységet és csökkentheti a folyamatok következetlenségből eredő veszteségeket. A digitális átalakulás szükségességét az Egan (UK, 1998) jelentés [33] is hangsúlyozza, amely kiemeli az integrált folyamatok és az együttműködésen alapuló megközelítés kulcsfontosságú szerepét az építőiparban.

Ennek eredményeképpen, míg az elmúlt 10 000 évben az adatkezelők fő problémája az adatok hiánya volt, az adatok és az adatkezelési rendszerek lavinájával a felhasználók és az adatkezelők egy problémával szembesülnek: az adatok túlburjánzsával, ami megnehezíti a jogilag helyes és minőségi információk megtalálását.

Az adatok széttagolt silói elkerülhetetlenül az adatminőség csökkenésének súlyos problémájához vezetnek. Több független rendszer esetén ugyanazok az adatok különböző verziókban létezhetnek, gyakran egymásnak ellentmondó értékekkel, ami további nehézségeket okoz a felhasználók számára, akiknek meg kell határozniuk, hogy melyik információ releváns és megbízható.

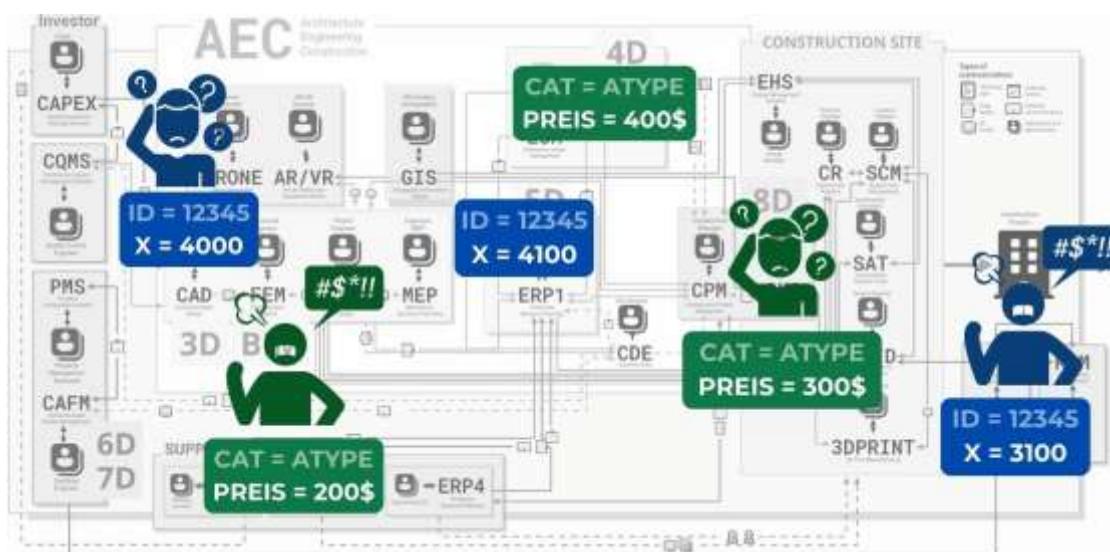
Duplikáció és az adatok minőségének hiánya a széttagoltság következményeként.

Az adatsilók problémája miatt a vezetőknek jelentős időt kell tölteniük az adatok keresésével és egyeztetésével. A minőségi problémák elleni védekezés érdekében a vállalatok összetett információkezelési struktúrákat hoznak létre, amelyekben a vezetők egy vertikális csoportja felelős az adatok kereséséért, ellenőrzéséért és egyeztetéséért. Ez a megközelítés azonban csak növeli a bürokráciát

és lassítja a döntéshozatalt. Minél több az adat, annál nehezebb elemezni és értelmezni, különösen, ha nincs egységes szabvány az adatok tárolására és feldolgozására.

A szoftveralkalmazások és rendszerek sokaságával, amelyek az elmúlt évtizedben úgy nőttek, mint gomba az eső után, a silók és a nem megfelelő adatminőség problémája egyre fontosabbá vált a végfelhasználók számára. Ugyanazok az adatok, de különböző értékekkel, ma már különböző rendszerekben és alkalmazásokban találhatók (2.1-6. ábra). Ez nehézségekhez vezet a végfelhasználók számára, amikor megpróbálják meghatározni, hogy a sok rendelkezésre álló adat közül melyik változat a releváns és helyes. Ez hibákhoz vezet az elemzésekben és végső soron a döntéshozatalban.

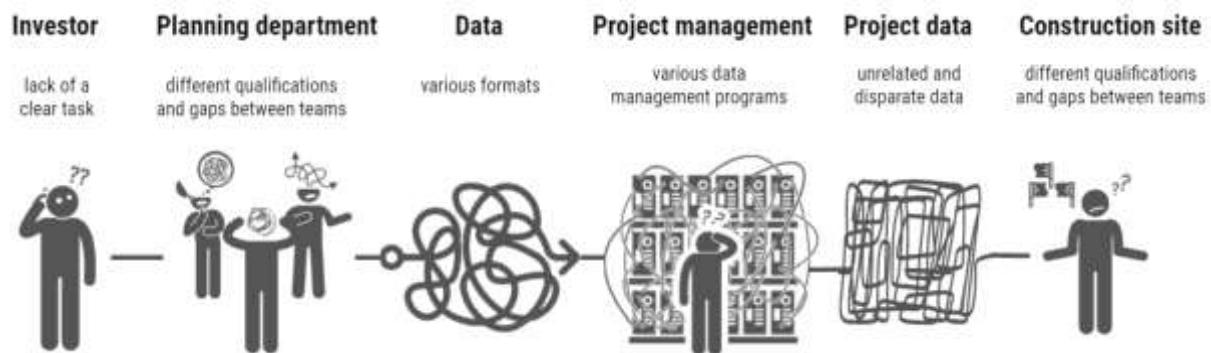
A megfelelő adatok megtalálásának problémái ellen a vállalatvezetők többszintű bürokráciát hoznak létre az ellenőrzési vezetőkből. Az Ő feladataik az, hogy gyorsan megtalálják, ellenőrizzék és elküldjék a szükséges adatokat táblázatok és jelentések formájában, eligazodva az eltérő rendszerek útvesztőjében.



Ábra 2.1-6 A megfelelő adatok megtalálására törekedve a vezetőknek biztosítaniuk kell az adatok minőségét és jogi megbízhatóságát a különböző rendszerek között.

A gyakorlatban azonban ez a modell új összetettséget teremt. Ha az adatokat kézzel kezelik, és az információk sok, egymástól független döntésben szétszóródnak, minden kísérlet, hogy a döntéshozók piramisán keresztül (2.1-7. ábra) pontos és naprakész információkhoz jussanak, szűk keresztmetszetté válik - időigényes és hibalehetőségekkel teli.

A helyzetet súlyosbítja a digitális megoldások lavinája. A szoftverpiacot továbbra is elárasztják az ígéretesnek tűnő új eszközök. De egyértelmű adatkezelési stratégia nélkül ezek a megoldások nem integrálódnak egységes rendszerbe, hanem a bonyolultság és a duplikáció további rétegeit hozzák létre. Ennek eredményeképpen a folyamatok egyszerűsítése helyett a vállalatok egy még széttöredezettebb és kaotikusabb információs környezetben találják magukat.



Ábra: 2.1-7 A rendszerek összetettsége és az adatformátumok sokfélesége a következetesség elvesztéséhez vezet az építési folyamat során.

Mindezek az eltérő megoldások sokaságának kezelésével kapcsolatos problémák előbb-utóbb egy fontos felismerésre késztetik a vállalatvezetést: nem az adatmennyiséggel vagy a következő "egyméretű" adatfeldolgozó eszköz keresésével van baj. Az igazi ok az adatok minőségében, valamint abban rejlik, hogy a szervezet hogyan hozza létre, fogadja, tárolja és használja azokat.

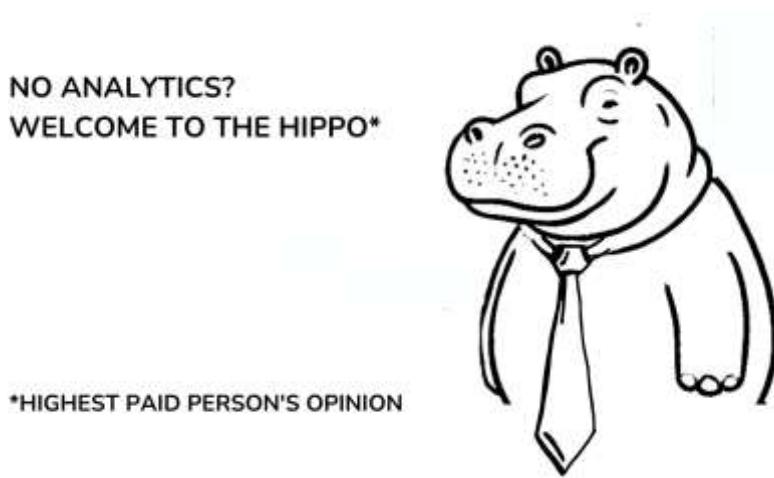
A fenntartható siker kulcsa nem az új "varázslatos" alkalmazások hajszolásában rejlik, hanem a vállalon belüli adatkultúra kiépítésében. Ez azt jelenti, hogy az adatokat stratégiai értékként kell kezelní, és az adatminőséget, az integritást és a relevanciát a szervezet minden szintjén prioritássá kell tenni.

A "minőség kontra mennyiség" dilemma megoldása egy olyan egységes adatszerkezet létrehozásában rejlik, amely kiküszöböli a duplikációkat, megszünteti a következetlenségeket és egységesíti az információáramlást. Ez a struktúra egyetlen, megbízható adatforrást biztosít, amely alapján megalapozott, pontos és időszerű döntéseket lehet hozni.

Ellenkező esetben, mint ahogy ez még mindig gyakran előfordul, a vállalatok továbbra is a HiPPO szakértőinek szubjektív véleményére és intuitív értékelésére hagyatkoznak, nem pedig megbízható tényekre. Az építőiparban, ahol a szakértelem hagyományosan jelentős szerepet játszik, ez különösen szembetűnő.

HiPPO vagy a vélemények veszélye a döntéshozatalban

Az építőiparban a legfontosabb döntéseket hagyományosan a tapasztalat és a szubjektív megítélés alapján hozzák meg. Időszerű és megbízható adatok nélkül a vállalatvezetők kénytelenek vakon cselekedni, az objektív tények helyett a legjobban fizetett alkalmazottak intuíciójára (HiPPO - Highest Paid Person's Opinion) hagyatkozva (2.1-8. ábra).



Ábra 2.1-8 Analitika hiányában az üzleti élet a tapasztalt szakemberek szubjektív véleményétől függ.

Ez a megközelítés indokolt lehet egy stabil és lassan változó környezetben, de a digitális átalakulás korában komoly kockázatot jelent. Az intuición és találkodáson alapuló döntések hajlamosak a torzításra, gyakran megalapozatlan hipotéziseken alapulnak, és nem veszik figyelembe az adatokban tükröződő összetett képet

Amit egy vállalat döntéshozatali szintjén intelligens vitának adnak ki, gyakran semmi konkréturnon nem alapul. Egy vállalat sikerének nem a szakértők tekintélyétől és fizetésétől kellene függenie, hanem attól, hogy képesek-e hatékonyan dolgozni az adatokkal, felismerni a mintákat és megalapozott döntéseket hozni.

Fontos, hogy felhagyunk azzal a gondolattal, hogy a tekintély vagy a tapasztalat automatikusan azt jelenti, hogy egy döntés helyes. Az adatvezérelt megközelítés megváltoztatja a játékszabályokat: a döntéshozatal alapját mostantól az adatok és az elemzések, nem pedig a pozíció és a fizetés képezik. A nagyméretű adatok, a gépi tanulás és a vizuális analitika lehetővé teszi számunkra, hogy mintákat azonosítsunk, és találkodások helyett tényekre támaszkodjunk (1.1-4. ábra).

Adatok nélkül csak egy újabb ember vagy, akinek véleménye van [34].

- W. Edwards Deming, tudós és vezetési tanácsadó

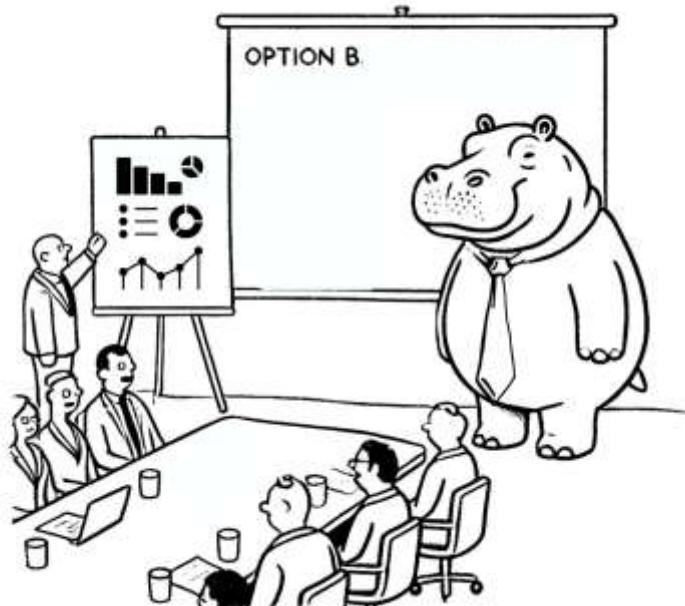
A modern adatkezelési módszerek a vállalaton belüli tudás folyamatosságát is biztosítják. A világosan leírt folyamatok, az automatizálás és a szisztematikus megközelítés lehetővé teszi, hogy még a kulcsszerepeket is át lehessen adni anélkül, hogy a hatékonyság csökkenne.

Az adatokba vetett vak bizalom azonban súlyos hibákhoz is vezethet. Az adatok önmagukban csak számok gyűjteménye. Megfelelő elemzés, kontextus és a minták azonosításának képessége nélkül nincs értékük, és nem képesek a folyamatok irányítására. A sikeres kulcsa nem a HiPPO intuíció és az analitika közötti választásban rejlik, hanem olyan intelligens eszközök kiépítésében, amelyek az eltérő információkat kezelhető, megalapozott döntésekkel alakítják át.

A digitális építési környezetben nem a ranglétra és a hierarchiában elfoglalt hely válik döntő sikertényezővé, hanem a reagálóképesség, a döntések pontossága és az erőforrások hatékonysága

Az adatok eszközök, nem pedig abszolút igazságok. Az adatoknak ki kell egészíteniük az emberi gondolkodást, nem pedig helyettesíteniük azt. Az analitika előnyei ellenére, az adatok nem helyettesíthetik teljesen az emberi intuíciót és tapasztalatot. Szerepük az, hogy segítsenek pontosabb és megalapozottabb döntéseket hozni.

A versenyelőny nem csupán a szabványoknak való megfeleléssel érhető el, hanem azzal, hogy a mindenki számára azonos erőforrások hatékony felhasználásában képesek vagyunk a versenytársak fölé kerekedni. A jövőben az adatkezelési készségek ugyanolyan fontosak lesznek, mint egykor az írástudás vagy a matematikai készségek. Az adatok elemzéséhez és értelmezéséhez értő szakemberek pontosabb döntéseket tudnak majd hozni, kiszorítva azokat, akik csak a személyes tapasztalataikra hagyatkoznak (2.1-9. ábra).



Ábra: 2.1-9 A döntéseknek objektív elemzésen kell alapulniuk, nem pedig a legmagasabb fizetésű alkalmazott véleményén.

A vezetők, szakemberek és mérnökök a adatelemzői lesznek, akik a projektek szerkezetét, dinamikáját és kulcsmutatóit tanulmányozzák. Az emberi erőforrások a rendszer elemeivé válnak, és a hatékonyság maximalizálása érdekében rugalmas adatvezérelt testreszabást igényelnek.

A nem megfelelő adatok használata esetén a hibák sokkal kisebbek, mint az adatok hiányában [35].

- Charles Babbage, az első analitikus számológép feltalálója

A nagyméretű adatok megjelenése és az LLM (Large Language Models) bevezetése nemcsak az elemzés módját, hanem a döntéshozatal természetét is gyökeresen megváltoztatta. Míg korábban az ok-okozati összefüggésekre (miért történt valami - diagnosztikus analitika) helyezték a hangsúlyt (1.1-4. ábra), ma már a jövő előrejelzésének képessége (prediktív analitika), a jövőben pedig a preskriptív analitika kerül előtérbe, ahol a gépi tanulás és az AI a legjobb választást javasolja a döntéshozatali folyamatban.

Az SAP™ "New Study Finds Nearly Half of Executives Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36] című új tanulmánya szerint a felsővezetők 44%-a hajlandó lenne megváltoztatni korábbi döntését az AI tanácsai alapján, és 38% bízna abban, hogy az AI meghozza helyettük az üzleti döntéseket. Eközben a vezetők 74%-a azt mondta, hogy jobban bízik az AI-tanácsokban, mint a barátaiban és a családjában, és 55%-uk olyan vállalatoknál dolgozik, ahol az AI-alapú meglátások helyettesítik vagy gyakran megkerülik a hagyományos döntéshozatali módszereket - különösen az 5 milliárd dollárnál nagyobb éves bevétellel rendelkező szervezeteknél. Emellett a válaszadók 48%-a napi szinten használja a generatív AI-eszközöket, ebből 15% naponta többször is.

Az LLM és az automatizált adatkezelési rendszerek fejlődésével új kihívás merül fel: hogyan lehet az információt hatékonyan felhasználni anélkül, hogy az értékét elveszítené az inkompatibilis formátumok és heterogén források káoszában, amelyet az üzleti folyamatok növekvő összetettsége és dinamikája egészít ki.

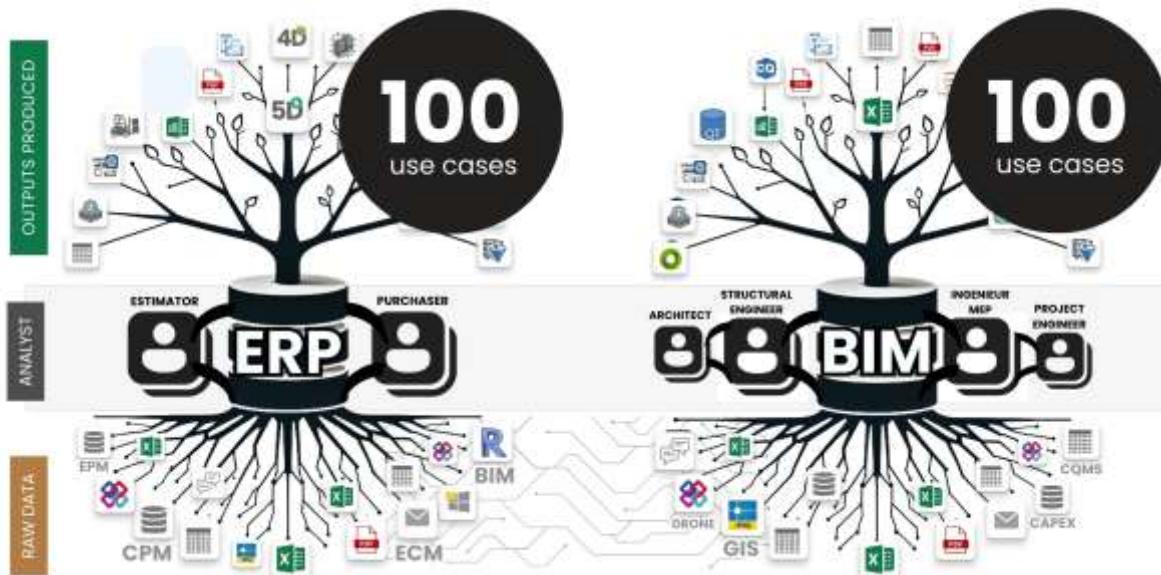
Az üzleti folyamatok összetettségének és dinamizmusának folyamatos növekedése

Az építőipar ma komoly kihívásokkal néz szembe az adat- és folyamatmenedzsment terén. A fő kihívást a széttöredezett információs rendszerek, a túlzott bürokrácia és a digitális eszközök közötti integráció hiánya jelenti. Ezek a kihívások fokozónak, ahogy maguk az üzleti folyamatok is egyre összetettebbé válnak - a technológia, a változó ügyfélkövetelmények és a fejlődő szabályozások hatására.

Az építési projektek egyedisége nemcsak műszaki sajátosságaikból, hanem a különböző országok nemzeti szabványai és szabályozási követelményei közötti különbségekből is adódik (4.2-10. ábra, 5.1-7. ábra). Ez minden egyes projekt esetében rugalmas, egyedi megközelítést igényel, ami a hagyományos moduláris ellenőrzési rendszerekben nehezen megvalósítható. A folyamatok összetettsége és a nagy adatmennyisége miatt sok vállalat fordul a speciális megoldásokat kínáló gyártókhöz. A piac azonban túlterhelt - sok startup kínál hasonló termékeket, amelyek szűk feladatokra összpontosítanak. Ennek eredményeként az adatkezelés holisztikus megközelítése gyakran elvész.

Az új technológiák és piaci követelmények folyamatos áramlásához való alkalmazkodás a versenyképesség kritikus tényezőjévé válik. A meglévő saját fejlesztésű alkalmazások és moduláris rendszerek azonban kevésbé alkalmazkodóképesek - bármilyen változtatás gyakran hosszadalmas és költséges felülvizsgálatot igényel a fejlesztők részéről, akik nem minden ismerik az építési folyamatok sajátosságait.

A vállalatok a technológiai lemaradás foglyai, és az új frissítésekre várnak, ahelyett, hogy azonnal innovatív integrált megközelítéseket alkalmaznának. Ennek eredményeképpen az építőipari szervezetek belső struktúrája gyakran egymással összekapcsolt hierarchikus és gyakran zárt rendszerek komplex ökoszisztémája, amelyet a vezetők többszintű hálózatán keresztül koordinálnak (2.1-10. ábra).

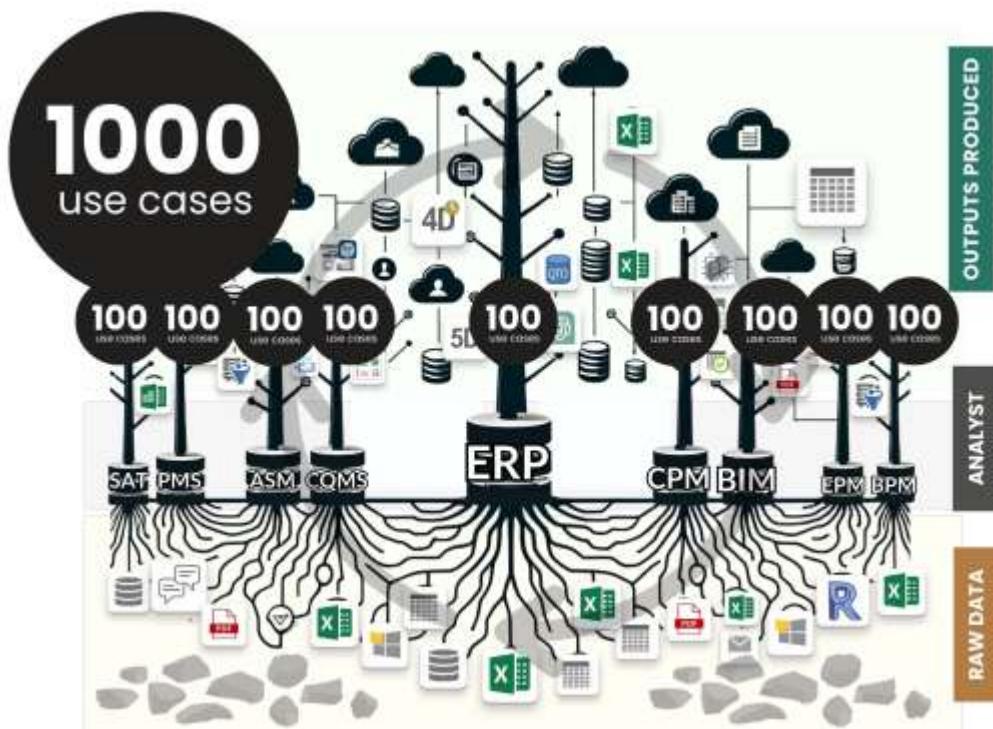


Ábra 2.1-10 A vállalatok egymással összekapcsolt rendszerekből állnak, amelyek összekapcsolódása automatizálást igénylő folyamatokat alkot.

A Kanadai Építőipari Szövetség és a KPMG Canada 2021 [37] által készített tanulmány szerint a vállalatoknak csupán 25%-a gondolja úgy, hogy a technológiai bevezetés vagy a digitális megoldások tekintetében jelentős vagy eltérő helyzetben van a versenytársakhoz képest. A válaszadók minden össze 23%-a számolt be arról, hogy megoldásai jelentősen vagy erősen adatvezéreltek. Ugyanakkor a felmérésben részt vevők többsége egy sor más technológia használatát pusztán kísérleti jellegűnek jellemzte, vagy elismerte, hogy egyáltalán nem használja azokat.

Ez a vonakodás a technológiai kísérletekben való részvételtől különösen a nagy infrastrukturális projektek esetében nyilvánvaló, ahol a hibák dollármilliókba kerülhetnek. Még a legfejlettebb technológiák - digitális ikrek, prediktív analitika - is gyakran nem hatékonyságuk, hanem a valós projektekben bizonyított megbízhatóságuk hiánya miatt ütköznek ellenállásba.

A Világgazdasági Fórum (WEF) "Az építőipar jövőjének alakítása" című jelentése [5] szerint az új technológiák bevezetése az építőiparban nemcsak technikai nehézségekkel, hanem az ügyfelek részéről pszichológiai akadályokkal is szembesül. [5] szerint az új technológiák bevezetése az építőiparban nemcsak technikai nehézségekkel, hanem a vásárlók részéről pszichológiai akadályokkal is szembesül. Sok megrendelő attól tart, hogy a fejlett megoldások alkalmazása a projektjeiket kísérleti helyszínné és "kísérleti nyulakká" teszi, a kiszámíthatatlan következmények pedig további költségeket és kockázatokat eredményezhetnek.



Ábra 2.1-11 A megoldások piaca minden egyes adatfelhasználási esethez kínál alkalmazásokat a folyamatok optimalizálására és automatizálására.

Az építőipar nagyon változatos: a különböző projektek különböző követelményekkel rendelkeznek, regionális sajátosságokkal, törvényi osztályozási előírásokkal (4.2-10. ábra), számítási szabványokkal (5.1-7. ábra) stb. Ezért gyakorlatilag lehetetlen olyan saját, univerzális alkalmazást vagy rendszert létrehozni, amely tökéletesen megfelelne mindenek követelményeinek és a projekt sajátosságainak.

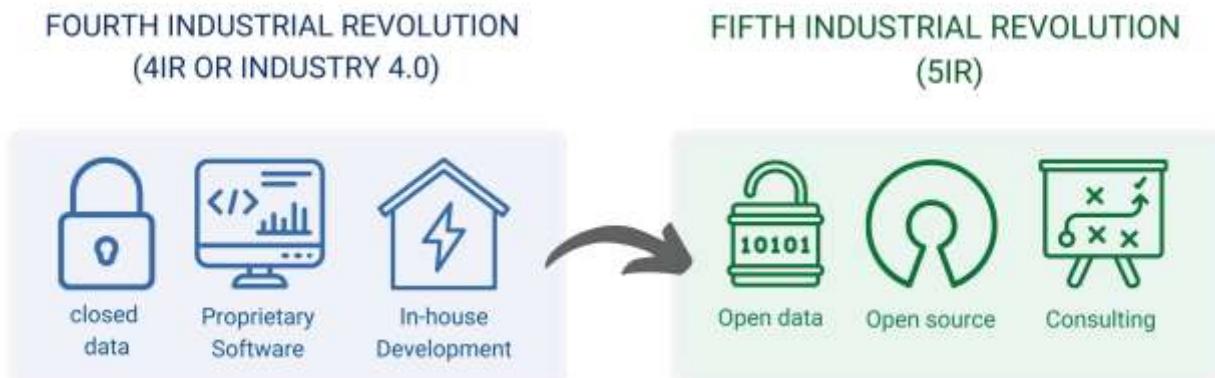
A rendszerek növekvő összetettségével és a szoftvergyártóktól való függőséggel való megbirkózás érdekében egyre inkább felismerik, hogy a hatékony adatkezelés kulcsa nemcsak a nyitottság és a szabványosítás, hanem magának a folyamatarchitektúrának az egyszerűsítése is. Az üzleti folyamatok növekvő összetettsége és dinamizmusa új megközelítéseket igényel, ahol az adatok felhalmozásáról az adatok strukturálására és szervezésére helyeződik át a hangsúly. Ez a váltás lesz a következő lépés az építőipar fejlődésében, amely a szoftvergyártók dominanciájának végét és az értelmes információszervezés korszakának kezdetét jelzi.

Az egyformán alkalmazható megoldások korlátainak felismerése és a növekvő komplexitással szembeni sebezhetőség a prioritások eltolódását eredményezi a zárt platformoktól és az adatok felhalmozásától az átláthatóság, az alkalmazkodóképesség és a strukturált információkezelés felé. Ez a gondolkodásbeli változás a globális gazdaság és technológia szélesebb körű változásait tükrözi, amelyeket az úgynevezett "ipari forradalmak" szemszögéből írunk le. Ahhoz, hogy megértsük, merre tart az építőipar és milyen irányt vesz a jövőben, meg kell vizsgálni az iparág helyét a negyedik és ötödik ipari forradalom kontextusában - az automatizálástól és a digitalizációtól a személyre szabásig, a nyílt szabványokig és a szolgáltatásalapú adatmodellig.

A negyedik ipari forradalom (Ipar 4.0) és az ötödik ipari forradalom (Ipar 5.0) az építőiparban.

A technológiai és gazdasági szakaszok olyan elméleti fogalmak, amelyeket a társadalom és a gazdaság fejlődésének leírására és elemzésére használnak a fejlődés különböző szakaszában. Ezeket a különböző kutatók és szakértők eltérően értelmezhetik.

- **A negyedik ipari forradalom** (4IR vagy Ipar 4.0) az információs technológiához, az automatizáláshoz, a digitalizációhoz és a globalizációhoz kapcsolódik. Ennek egyik kulcseleme a saját szoftvermegoldások, azaz speciális feladatokra és vállalatokra tervezett speciális digitális termékek létrehozása. Ezek a megoldások gyakran az IT-infrastruktúra fontos részévé válnak, de további módosítások nélkül nehezen skálázhatók.
- **Az ötödik ipari forradalom** (5IR) jelenleg a koncepcióalkotás és a fejlesztés korábbi szakaszában van, mint a 4IR. Alapelvei közé tartozik a termékek és szolgáltatások fokozott személyre szabása. Az 5IR az alkalmazkodóképesebb, rugalmasabb és személyre szabottabb gazdasági tevékenység irányába mutató mozgalom, amelynek középpontjában a személyre szabás, a tanácsadás és a szolgáltatásorientált modellek állnak. Az ötödik gazdasági mód egyik fő szempontja az adatok felhasználása a döntéshozatalban, ami gyakorlatilag lehetetlen a nyílt adatok és nyílt eszközök használata nélkül (2.1-12. ábra).



Ábra 2.1-12 A negyedik minta a megoldásokra, míg az ötödik minta a személyre szabásra és az adatokra összpontosít.

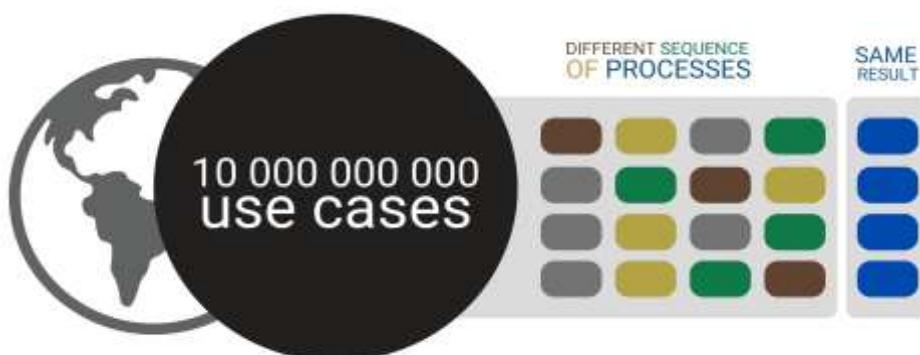
Egy alkalmazás létrehozása az építőiparban működő vállalatok számára tíz vagy száz szervezet számára nem garantálja, hogy jelentős módosítások és fejlesztések nélkül sikeresen átültethető más vállalatokra, régióra vagy országokra. Az ilyen megoldások sikeres skálázásának valószínűsége továbbra is alacsony, mivel minden egyes szervezet egyedi folyamatokkal, követelményekkel és feltételekkel rendelkezik, amelyek személyre szabott alkalmazásokat igényelhetnek.

Fontos megérteni, hogy a technológiai megoldások sikeres integrációja már ma is mélyen személyre szabott megközelítést jelent minden egyes folyamathoz, projekthez és vállalathoz. Ez azt jelenti, hogy még egy univerzális keretrendszer, eszköz vagy program kifejlesztése után is részletes alkalmazkodást és testre szabást igényel, hogy megfeleljen az egyes vállalatok és projektek egyedi követelményeinek és feltételeinek.

A PwC "Az ötödik ipari forradalom dekódolása" című jelentése szerint [38] a különböző iparágak felsővezetőinek mintegy 50%-a idén a fejlett technológia és az emberi szakértelem integrációjára támaszkodik. Ez a megközelítés lehetővé teszi számukra, hogy gyorsan alkalmazkodjanak a terméktervezés vagy a vevői igények változásaihoz, személyre szabott termelést létrehozva.

Minden egyes folyamat egyedi funkció vagy alkalmazás kifejlesztését igényli, ami a globális építőipar méretét és a projektek sokféleségét tekintve rengeteg üzleti eset létezéséhez vezet, amelyek minden alkalommal egyedi Pipeline logikát képviselnek (2.1-13. ábra). minden ilyen esetnek megvannak a maga sajátosságai, és egyedi megközelítést igényel. Ugyanazon analitikai probléma lehetséges megoldásainak sokféleségét a különböző megközelítések keretében a gépi tanulásnak és a Titanic-adatkészlet elemzésének szentelt fejezetben (9.2-9. ábra) fogjuk részletebben megvizsgálni.

A csővezeték a digitális folyamatokkal összefüggésben olyan tevékenységek, folyamatok és eszközök sorozata, amelyek lehetővé teszik az adatok és a munka automatizált vagy strukturált áramlását a projekt életciklusának különböző szakaszain keresztül.



Ábra 2.1-13 Az üzleti esetek egyedisége és változékonysága lehetetlenné teszi a skálázható zárt platformok és eszközök létrehozására tett kísérleteket.

A digitális átalakulás hatására már sok tekintetben megváltozott az életünk, és ma már az építőipar

gazdasági fejlődésének új szakaszáról beszélhetünk. Ebben az "új gazdaságban" a verseny más szabályok szerint fog szerveződni: aki képes a nyilvános tudást és a nyílt adatokat hatékonyan keresett termékekkel és szolgáltatásokkal alakítani, az kulcsfontosságú előnyre tesz szert az ötödik ipari forradalom körülményei között.

Ahogy Kate Maskus közgazdász megjegyzi a "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39] című könyvében. 2012 [39], "globális tudásalapú gazdaságban élünk, és a jövő azoké, akik tudják, hogyan kell a tudományos felfedezéseket árucikké tenni".

Az ötödik gazdasági módra való áttérés a zárt informatikai megoldásokról a nyílt szabványok és platformok felé való elmozdulást jelenti. A vállalatok kezdenek majd eltávolodni a hagyományos szoftvertermékektől a szolgáltatásorientált modellek javára, ahol a fő eszköz nem a szabadalmaztatott technológiák, hanem az adatok lesznek.

A Harvard Business School 2024 tanulmánya [40] kimutatja a nyílt forráskódú szoftverek (Open Source Software, OSS) hatalmas gazdasági értékét. A tanulmány szerint az OSS az összes szoftverkód 96%-ában jelen van, és egyes kereskedelmi szoftverek 99,9%-ban OSS komponensekből állnak. Az OSS nélkül a vállalatok 3,5-szer többet költenének szoftverre.

A vállalati ökoszisztémák kiépítése a globális trendeket követve fokozatosan át fog térní egy ötödik gazdasági paradigmára, ahol az adatközpontú elemzési és tanácsadói szolgáltatások nagyobb prioritást kapnak, mint a mereven meghatározott felhasználási forgatókönyvekkel rendelkező elszigetelt, zárt megoldások.

A digitalizáció kora megváltoztatja az erőviszonyokat az iparágban: ahelyett, hogy a szállítók megoldásaira támaszkodnának, a vállalatok versenyképességüket az adatok hatékony felhasználására való képességükre alapozzák. Ennek eredményeképpen az építőipar az örökölt merev rendszerektől a rugalmas, alkalmazkodó ökoszisztémák felé mozdul el, ahol a nyílt szabványok és az interoperabilis eszközök képezik majd a projektmenedzsment alapját. Az alkalmazásgyártók dominanciájának vége egy olyan új környezetet teremt, amelyben az értéket nem a zárt forráskód és a speciális csatlakozók birtoklása, hanem az adatok stratégiai előnyé alakításának képessége határozza meg.



FEJEZET 2.2.

A KÁOSZ RENDDÉ ALAKÍTÁSA ÉS A KOMPLEXITÁS CSÖKKENTÉSE

A redundáns kód és a zárt rendszerek mint a termelékenység javításának akadályai

Az elmúlt évtizedekben az informatikai szféra technológiai változásait elsősorban a szoftvergyártók hajtották. Ők szabták meg a fejlődés irányát, ők határozták meg, hogy a vállalatok mely technológiákat vegyék át, és melyeket hagyják el. Abban a korszakban, amikor a silázott megoldásokról a központosított adatbázisokra és integrált rendszerekre való áttérés következett be, a gyártók licencelt termékeket népszerűsítettek, amelyek a hozzáférés és a skálázhatóság ellenőrzését biztosították. Később, a felhőtechnológiák és a Software as a Service (SaaS) modellek megjelenésével ez az ellenőrzés előfizetési modellé alakult, ami a felhasználókat a digitális szolgáltatások hűséges ügyfeleivé tette.

Ez a megközelítés paradoxonhoz vezetett: a létrehozott programkódok soha nem látott mennyisége ellenére csak egy kis részét használják fel ténylegesen. Talán százszer vagy ezerszer több kód van a szükségesnél, mert ugyanazokat az üzleti folyamatokat tucatnyi vagy száz programban írják le és másolják meg különböző módon, még ugyanazon a vállalon belül is. Ugyanakkor a fejlesztési költségeket már kifizették, és ezek a költségek nem térülnek meg. Ennek ellenére az ipar továbbra is ezt a ciklust folytatja, és a végfelhasználó számára minimális hozzáadott értékkal rendelkező új termékeket hoz létre, gyakrabban a piaci elvárások, mint a valós igények nyomása alatt.

A Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41] szerint a szoftverfejlesztés költségei számos tényezőtől, többek között a rendszer összetettségétől és a választott technológiától függően jelentősen eltérhetnek. Történelmileg a fejlesztési költségek 2008-ban forráskódosoroknál (SLOC) 100 dollár körül mozogtak, míg a karbantartási költségek SLOC-onként akár 4000 dollárra is emelkedhetnek.

A CAD-alkalmazások egyetlen összetevője - a geometriai mag - több tízmillió sornyi kódot tartalmazhat (6.1-5. ábra). Hasonló helyzet figyelhető meg az ERP-rendszereknél (5.4-4. ábra), amelyek komplexitásának tárgyalására a könyv ötödik részében térünk vissza. Közelebbről megvizsgálva azonban kiderül, hogy e kód nagy része nem ad hozzáadott értéket, hanem csupán "postásként" működik - mechanikusan mozgatja az adatokat az adatbázis, az API, a felhasználói felület és a rendszer egyéb táblái között. Az úgynevezett üzleti logika kritikus fontosságáról szóló népszerű mítosz ellenére a rideg valóság sokkal prózaibb: a modern kódállományok tele vannak elavult sablonblokkokkal (örökolt kód), amelyeknek egyetlen célja a táblák és komponensek közötti adatátvitel biztosítása anélkül, hogy befolyásolnák a döntéshozatal vagy az üzleti hatékonyságot.

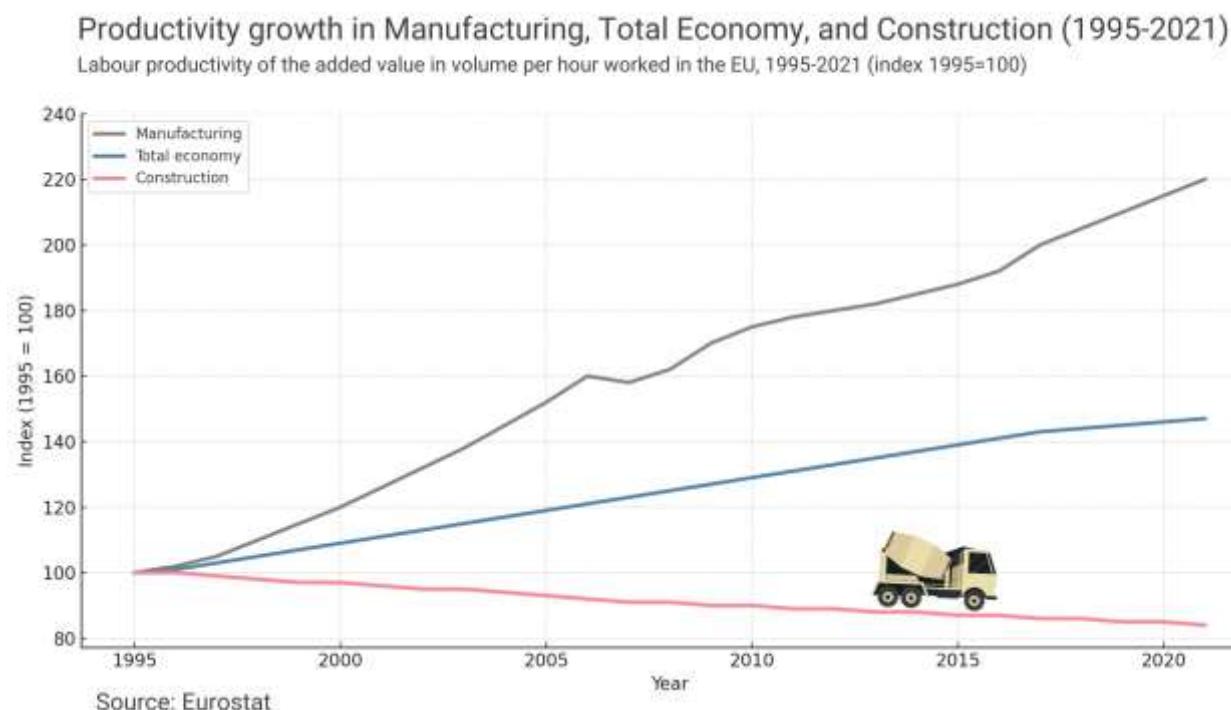
Ennek eredményeképpen a különböző forrásokból származó adatokat feldolgozó zárt megoldások elkerülhetetlenül zavaros "spagetti ökoszisztemákká" válnak. Ezeket az összetett, egymásba fonódó rendszereket csak egy seregni, félig rutinszerűen dolgozó menedzser tudja kezelní. Az adatkezelés ilyen megszervezése nemcsak az erőforrások szempontjából nem hatékony, hanem kritikus sebezhetőségeket is teremt az üzleti folyamatokban, és a vállalatot azon szakemberek szűk körétől teszi függővé, akik értik, hogyan működik ez a technológiai útvesztő.

A kódmennyiség, az alkalmazások számának folyamatos növekedése és a gyártók által kínált koncepciók egyre összetettebbé válása természetes eredményhez vezetett - az építőipari IT-ökoszisztemáma

összetettségének növekedéséhez. Ez hatástellanná tette a digitalizáció gyakorlati megvalósítását az alkalmazások számának növelésével az ágazatban. A felhasználói igények kellő figyelembevétele nélkül létrehozott szoftvertermékek gyakran jelentős erőforrásokat igényelnek a bevezetéshez és a támogatáshoz, de nem hozzák meg a várt hozamot.

A McKinsey "Az építőipari termelékenység növelése" című tanulmánya szerint [42] szerint az elmúlt két évtizedben a globális munkatermelékenység növekedése az építőiparban átlagosan csak évi 1% volt, szemben a világgazdaság egészének 2,8%-os és a feldolgozóipar 3,6%-os növekedésével. Az Egyesült Államokban az egy dolgozóra jutó építőipari munkatermelékenység az 1960-as évek óta a felére csökkent [43].

A rendszerek növekvő összetettsége, az elszigeteltség és a zárt adatok akadályozzák a szakemberek közötti kommunikációt, ami az építőipart a legkevésbé hatékonyak közé teszi (2.2-1. ábra). 2040-re 22 millió dollárra emelkedik, ami jelentős hatékonyságnövelést igényel.



2.2-1. ábra A zárt és összetett adatok és ennek következtében a szakemberek közötti gyenge kommunikáció az építőipart a gazdaság egyik legkevésbé hatékony ágazatává tette ([44], [45] alapján).

Amint azt a McKinsey (2024) "Az építőipari termelékenység biztosítása többé nem választható" című tanulmánya is hangsúlyozza, az erőforrások növekvő szükössége és az ágazat növekedési ütemének megduplázására irányuló törekvése miatt az építőipar többé nem engedheti meg magának, hogy a jelenlegi termelékenységi szinteken maradjon [44]. Az előrejelzések szerint az építőipar globális költségei a 2023-as 13 millió dolláról az évtized végére jóval magasabb szintre emelkednek, ami a hatékonyság kérdését nemcsak aktuálissá, hanem kritikussá teszi.

A hatékonyság javításának egyik legfontosabb módja az alkalmazási struktúrák és az adatökoszisztemák-architektúrák elkerülhetetlen egységesítése és egyszerűsítése lesz. Ez a rationalizálási megközelítés

megszünteti a vállalati rendszerekben az évek során felhalmozódott felesleges absztrakciós rétegeket és szükségtelen komplexitást.

A silóktól az egységes adattárházig

Minél több adatot halmoz fel egy szervezet, annál nehezebb lesz valódi értéket kinyerni belőle. Az elszigetelt silókban tárolt információk széttöredezett jellege miatt a modern vállalatok üzleti folyamatai olyanok, mintha az építők ezernyi különböző raktárban tárolt anyagból próbálnának felhőkarcolót építeni. Az információtöbblet nemcsak a jogilag releváns információkhoz való hozzáférést nehezíti meg, hanem a döntéshozatalt is lassítja: minden lépést többször ellenőrizni és megerősíteni kell.

Minden feladat vagy folyamat külön táblázatba vagy adatbázisba van kötve, és a rendszerek közötti adatcsere bonyolult integrációkat igényel. Az egyik rendszerben felmerülő hibák és következetlenségek láncbábkat okozhatnak a többi rendszerben. A hibás értékek, a késedelmes frissítések és a duplikált információk arra kényszerítik az alkalmazottakat, hogy jelentős időt töltsenek manuálisan az adatok egyeztetésével és egyeztetésével. Ennek eredményeként a szervezet több időt tölt a széttagoltság következményeinek kezelésével, mint a folyamatok fejlesztésével és optimalizálásával

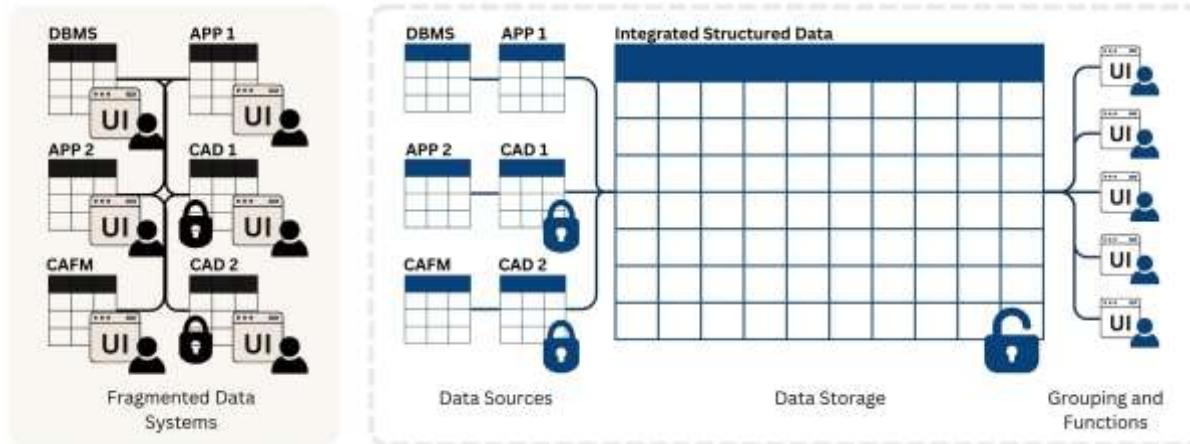
Ez a probléma egyetemes: egyes vállalatok továbbra is a káoszal küzdenek, míg mások az integrációban - az információáramlások központi tárolórendszerbe történő áthelyezésében - találnak megoldást. Gondoljon erre úgy, mint egy nagy táblára, ahol a feladatokhoz, projektekhöz és objektumokhoz kapcsolódó bármely entitást tárolhat. A tucatnyi különböző táblázat és formátum helyett egyetlen összefüggő adattár (2.2-2-2. ábra) jön létre, amely lehetővé teszi:

- minimalizálja az adatvesztést;
- kiküszöböli az információk folyamatos harmonizálásának szükségességét;
- az adatok elérhetőségének és minőségének javítása;
- az analitikus feldolgozás és a gépi tanulás egyszerűsítése

Az adatok közös szabványhoz való igazítása azt jelenti, hogy az információ a forrástól függetlenül egységes és géppel olvasható formátumba kerül. Az adatok ilyen szervezése lehetővé teszi azok integritásának ellenőrzését, valós idejű elemzését és azonnali felhasználását vezetői döntések meghozatalához.

Az integrált tárolórendszerek koncepcióját, valamint az analitikában és a gépi tanulásban való alkalmazásukat részletesebben a "Big Data Storage and Machine Learning" című fejezetben tárgyaljuk. Az adatmodellezés és -strukturálás témaörök részletesen az "Az adatok strukturált formába történő átalakítása" és a "Hogyan változtatják meg a szabványok a játékot: a véletlenszerű fájloktól a kidolgozott

adatmodellig" című fejezetek tárgyalják.



Ábra: 2.2-2 Az adatintegráció megszünteti a silókat, javítja az információk elérhetőségét és optimalizálja az üzleti folyamatokat.

Miután az adatok strukturálása és összevonása megtörtént, a következő logikus lépés az adatok hitelesítése. Egyetlen integrált adattárral ez a folyamat jelentősen leegyszerűsödik: nincs több ellentmondásos séma, duplikált struktúrák és bonyolult kapcsolatok a táblák között. minden információ egyetlen adatmodellhez igazodik, ami kiküszöböli a belső ellentmondásokat és felgyorsítja az érvényesítési folyamatot. A validálása és az adatminőség biztosítása minden üzleti folyamat sarkalatos pontja, és a könyv vonatkozó fejezetében részletesen foglalkozunk velük.

A végső szakaszban az adatokat csoportosítják, szűrik és elemzik. Különböző funkciókat alkalmaznak rájuk: aggregálás (összeadás, szorzás), táblázatok, oszlopok vagy sorok közötti számítások (2.2-4. ábra). Az adatokkal való munka lépések sorozatává válik: gyűjtés, strukturálás, validálás, átalakítás, analitikus feldolgozás és a végső alkalmazásokba való átvitel, ahol az információkat gyakorlati problémák megoldására használják fel. Az ilyen forgatókönyvek felépítéséről, a lépések automatizálásáról és a feldolgozási folyamatok felépítéséről bővebben az ETL-folyamatok és az adatvezetési megközelítés fejezetekben tárgyalunk.

A digitális átalakulás tehát nem csupán az információk kezelésének egyszerűsítéséről szól. Hanem az adatkezelés túlzott bonyolultságának megszüntetéséről, a káoszból a kiszámíthatóság, a több rendszerből a kezelhető folyamatok felé való elmozdulásról. Minél alacsonyabb az architektúra összetettsége, annál kevesebb kódra van szükség a támogatásához. A jövőben pedig a kód mint olyan teljesen eltűnhet, átadva helyét az intelligens ügynököknek, amelyek önállóan elemzik, rendszerezik és átalakítják az adatokat.

Az integrált tárolórendszerek lehetővé teszik az átállást az AI ügynökökre.

Minél kevésbé összetettek az adatok és a rendszerek, annál kevesebb kódot kell írni és karbantartani. A fejlesztés megtakarításának legegyszerűbb módja pedig az, ha teljesen megszabadulunk a kódtól, és azt adatokkal helyettesítjük. Amikor az alkalmazáskód-fejlesztés a kódról az adatmodellekre tér át, elkerülhetetlenül elmozdulás következik be az adatközpontú (adatvezérelt) megközelítés felé, mivel e fogalmak mögött teljesen más gondolkodásmód áll.

Ha valaki úgy dönt, hogy az adatokkal dolgozik a középpontban, másképp látja az adatok szerepét. Az adatok már nem csupán az alkalmazások "nyersanyagai" - mostantól az adatok jelentik az alapot, amelyre az architektúra, a logika és az interakció épül.

Az adatkezelés hagyományos megközelítése általában az alkalmazás szintjén kezdődik, és az építésben egy nehézkes bürokratikus rendszerre hasonlít: többszintű jóváhagyások, kézi ellenőrzések, a dokumentumok végtelen számú verziója a megfelelő szoftvertermékeken keresztül. A digitális technológiák fejlődésével egyre több vállalat lesz kénytelen áttérni a minimalizmus elvére - csak azt tárolni és használni, ami valóban szükséges, és amit használni is fognak.

A minimalizálás logikáját az eladók is átvették. Az adattárolás és -feldolgozás egyszerűsítése érdekében a felhasználói munka az offline alkalmazásokról és eszközökről a felhőszolgáltatásokba és az úgynevezett SaaS megoldásokba kerül át.

A SaaS koncepció (Software as a Service, azaz "szoftver mint szolgáltatás") a modern IT-infrastruktúrák egyik legfontosabb trendje, amely lehetővé teszi a felhasználók számára, hogy az interneten keresztül férjenek hozzá az alkalmazásokhoz anélkül, hogy saját számítógépéükön kellene telepíteniük és karbantartaniuk a szoftvereket.

A SaaS egyrészt megkönnyítette a skálázást, a verzióvezérlést és csökkentette a támogatási és karbantartási költségeket, másrészt azonban az adott alkalmazás logikájától való függés mellett a felhasználót teljesen függővé tette a szolgáltató felhő-infrastruktúrájától is. Ha egy szolgáltatás kiesik, az adatokhoz és az üzleti folyamatokhoz való hozzáférés átmenetileg vagy akár véglegesen is blokkolódhat. Ráadásul a SaaS-alkalmazásokkal való munka során a felhasználó összes adatát a szolgáltató szerverein tárolják, ami biztonsági és jogszabályi megfelelési kockázatot jelent. A tarifák vagy a felhasználási feltételek változása szintén megnövekedett költségeket vagy sürgős migráció szükségességét eredményezheti.

Az AI, az LLM -ügynökök és az adatközpontú megközelítés fejlődése megkérdőjelezte az alkalmazások jövőjét hagyományos formájukban és a SaaS vérehajtásában. Míg korábban az alkalmazásoknak és a szolgáltatásoknak az üzleti logikát kellett kezelnük és az adatokat feldolgozniuk, az AI-ügynökök megjelenésével ezek a funkciók átkerülhetnek az intelligens rendszerekhez, amelyek közvetlenül az adatokkal dolgoznak.

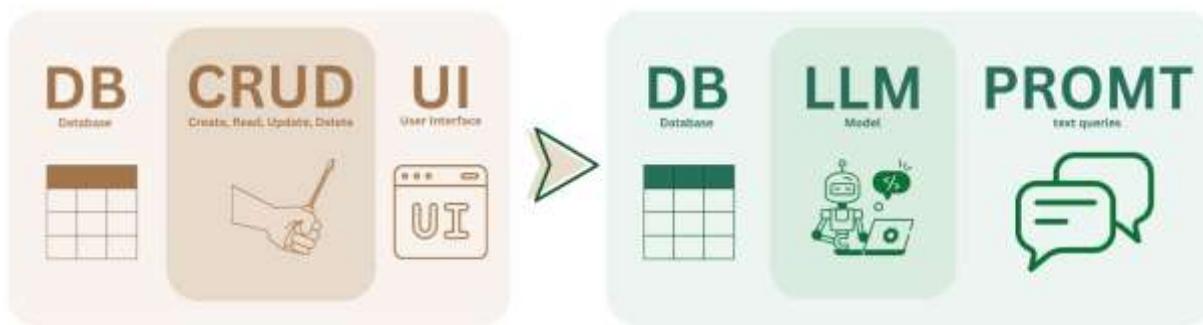
Ezért az informatikai részlegeken és a vezetőség szintjén egyre gyakrabban beszélnek hibrid architektúrákról, ahol az AI -ügynökök és a helyben elérhető megoldások kiegészítik a felhőszolgáltatásokat, csökkentve a SaaS -platformoktól való függőséget.

Az általunk alkalmazott megközelítés felismeri, hogy a hagyományos üzleti alkalmazások vagy SaaS -alkalmazások drámaian megváltozhatnak az ügynökök korában. Ezek az alkalmazások lényegében CRUD [create, read, update and delete] adatbázisok üzleti logikával. A jövőben azonban ezt a logikát a mesterséges intelligencia-ügynökök veszik át [46].

- Satya Nadella, a Microsoft vezérigazgatója, 2024.

Az adatközpontú megközelítés és az AI/LLM-ügynökök használata csökkentheti a felesleges folyamatokat, és ezáltal az alkalmazottak munkaterhelését. Ha az adatok megfelelően vannak rendszerezve, könnyebben válik az elemzés, a vizualizálás és a döntéshozatalban való alkalmazás. A végtelen jelentések és ellenőrzések helyett a szakemberek néhány kattintással vagy az LLM ügynökök segítségével automatikusan, kész dokumentumok és dashboardok formájában jutnak hozzá a naprakész információkhoz.

Az adatmanipulációban a mesterséges intelligencia eszközei (AI) és az LLM csevegések segítenek majd minket. Az utóbbi években a hagyományos CRUD-műveletek (create, read, update, delete) helyett a nagy nyelvi modellek (LLM) adatkezelésre való felhasználása felé indult el a trend. Az LLM-ek képesek a termesztes nyelv értelmezésére és a megfelelő adatbázis-lekérdezések automatikus generálására, ami leegyszerűsíti az adatkezelési rendszerekkel való interakciót (2.2-3. ábra).



Ábra 2.2-3 A mesterséges intelligencia felváltja és integrálja a tárolási és adatbázis-megoldásokat, fokozatosan kiszorítva a hagyományos alkalmazásokat és CRUD -műveleteket.

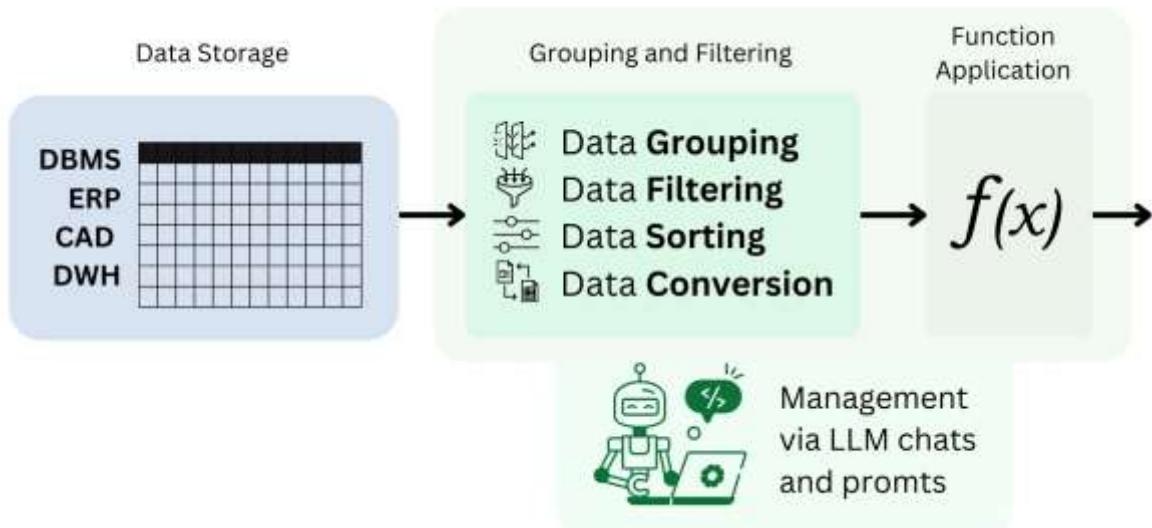
A következő 3-6 hónapban a mesterséges intelligencia a kód 90%-át írja majd, 12 hónapon belül pedig szinte az összes kódot mesterséges intelligencia generálhatja [47].

- Dario Amodei, az LLM vezérigazgatója Anthropic, 2025. március.

A mesterséges intelligencia fejlesztésére szolgáló eszközök (pl. GitHub Copilot) gyors fejlődése ellenére 2025-ben a fejlesztők még mindig kulcsszerepet játszanak ebben a folyamatban. Az AI-ügynökök egyre hasznosabb asszisztensekké válnak automatikusan értelmezik a felhasználó lekérdezéseit, SQL- és Pandas-lekérdezésekkel generálnak (erről bővebben a következő fejezetekben), vagy kódot írnak az adatok elemzéséhez. Ily módon a mesterséges intelligencia fokozatosan felváltja a hagyományos alkalmazás felhasználói felületeket.

A mesterséges intelligencia modellek, például a nyelvi modellek elterjedése a hibrid architektúrák fejlesztését fogja ösztönözni. Ahelyett, hogy teljesen lemondanánk a felhőmegoldásokról és a SaaS termékekről, a felhőszolgáltatások és a helyi adatkezelési rendszerek integrációjának lehetünk tanúi. A föderált tanulás például nagy teljesítményű AI modelleket tesz lehetővé anélkül, hogy az érzékeny adatokat a felhőbe kellene áthelyezni. Ily módon a vállalatok megtarthatják az adataik feletti ellenőrzést, miközben

hozzáférhetnek a fejlett technológiákhoz.



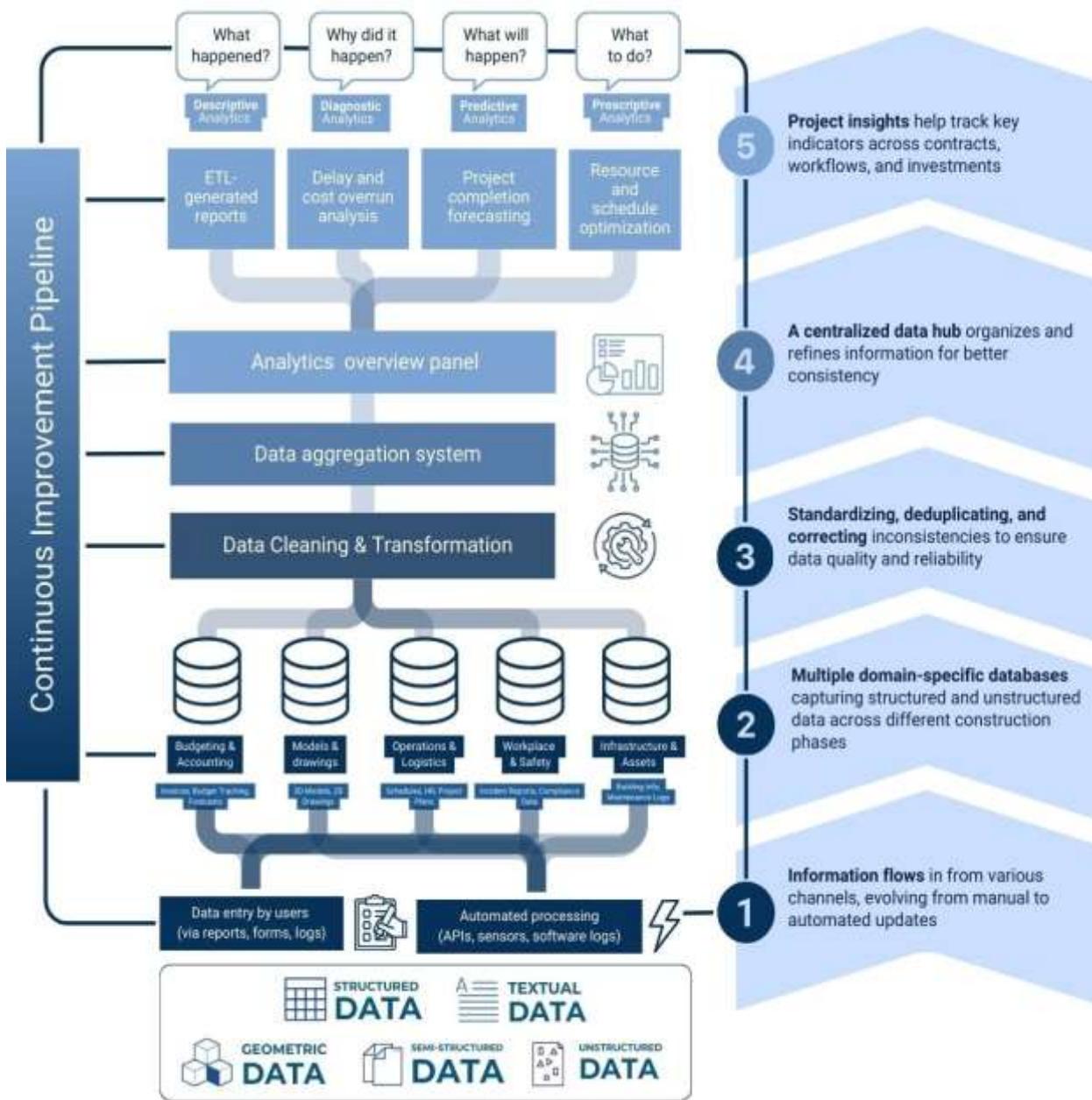
2.2-4 ábra A csoportosítás, szűrés és rendezés alapvető műveleteit, majd a függvények alkalmazását az LLM csevegések kezelik.

Az építőipar jövője a helyhez kötött megoldások, a felhőalapú teljesítmény és az intelligens modellek kombinációján alapul majd, amelyek együttesen hatékony és biztonságos adatkezelési rendszereket hoznak létre. Az LLM lehetővé teszi majd a mély műszaki ismeretekkel nem rendelkező felhasználók számára, hogy természetes nyelven megfogalmazott lekérdezésekkel interakcióba lépjenek az adatbázisokkal és adattárházakkal. Az LLM és az AI ügynökökről és működésükön bővebben az "LLM ügynökök és strukturált adatformátumok" című fejezetben beszélünk.

A megfelelően szervezett adatok és az egyszerű, könnyen használható LLM-alapú elemzőszközök nemcsak az információkkal való munkát könnyítik meg, hanem segítenek a hibák minimalizálásában, a hatékonyság növelésében és a folyamatok automatizálásában is.

Az adatgyűjtéstől a döntéshozatalig: az automatizáláshoz vezető út

A könyv későbbi részeiben részletesen megvizsgáljuk, hogyan lépnek kapcsolatba egymással a szakemberek, és hogyan válnak az adatok a döntéshozatal, az automatizálás és a működési hatékonyság alapjává. A 2.2-5. ábra egy példadiagramot mutat be, amely az adatfeldolgozási lépések sorrendjét mutatja be egy adatközpontú megközelítésben. Ez az ábra a folyamatos fejlesztési csővezetéket) szemlélteti, amelynek egyes részeit a könyv későbbi részében részletesen tárgyaljuk.



ábra 2.2-5 Egy példa a folyamatos adatjavító csővezetékre: az adatfeldolgozás és elemzés folyamata az építési projektekben.

Egy középvállalat üzleti folyamatait leíró rendszer többszintű elv alapján épül fel. Magában foglalja: az adatgyűjtést, tisztítást, aggregálást, analitikus feldolgozást és az eredmények alapján történő döntéshozatalt. Mindezeket a szakaszokat a könyv későbbi részében tanulmányozzuk - mind a elméleti kontextusban, mind gyakorlati példákon keresztül:

- Az első szinten **az adatbevitelre kerül sor** (3.1-1. ábra). Az információk kézzel (jelentéseken, ürlapokon, naplókon keresztül) és automatizált formában (API-ból, érzékelőkből, szoftverrendszerkből) egyaránt érkeznek. Az adatok különböző szerkezetűek lehetnek: geometriai, szöveges, strukturálatlan. Ebben a szakaszban szükség van az információáramlás

szabványosítására, strukturálására és egységesítésére.

- A következő szint az **adatfeldolgozás és -átalakítás**. Ez magában foglalja a tisztítási folyamatokat, a duplikációk eltávolítását, a hibák kijavítását és az információk további elemzésre való előkészítését (4.2-5. ábra). Ez a szakasz kritikus jelentőségű, mivel az analitika minősége közvetlenül az adatok tisztaságától és pontosságától függ.
- Az adatok ezután **speciális táblázatokba, adatkeretekbe vagy adatbázisokba kerülnek**, amelyek funkcionális területek szerint vannak felosztva: költségvetés és könyvelés, modellek és rajzok, logisztika, biztonság és infrastruktúra. Ez a felosztás lehetővé teszi az információk könnyű hozzáférését és keresztelemzését.
- Az adatokat ezután **összesítik és egy analitikus műszerfalon** (showcase) **jelenítik meg**. Itt leíró, diagnosztikai, prediktív és preskriptív elemzési módszereket alkalmaznak. Ez segít megválaszolni a kulcskérdéseket (1.1-4. ábra): mi történt, miért történt, mi fog történni a jövőben, és milyen intézkedéseket kell tenni. A rendszer például képes azonosítani a késedelmeket, megjósolni a projektek befejezését vagy optimalizálni az erőforrásokat.
- Végül az utolsó szint **elemző következetésteket és kulcsmutatókat** generál, amelyek segítenek a szerződések teljesítésének nyomon követésében, a beruházások kezelésében és az üzleti folyamatok javításában (7.4-2. ábra). Ezek az információk a döntéshozatal és a vállalat fejlesztési stratégiájának alapjává válnak.

Hasonlóképpen, az adatok a gyűjtésből a stratégiai irányításban való felhasználásig jutnak el. A könyv következő részeiben részletesen megvizsgáljuk az egyes szakaszokat, az adattípusokra, az adatfeldolgozási technikákra, az analitikai eszközökre és az építőiparban alkalmazott valós esetekre összpontosítva.

Következő lépések: a káoszból kezelhető rendszerré alakítása

Ebben a részben az információs silókkal kapcsolatos kihívásokat vizsgáltuk, és a túlzott rendszerkomplexitásnak az üzleti teljesítményre gyakorolt hatását vizsgáltuk, elemezve a negyedik ipari forradalomról az ötödikbe való átmenetet, ahol az alkalmazások helyett az adatok állnak a középpontban. Láttuk, hogy a silázott információs rendszerek hogyan gátolják a tudásmegosztást, és az informatikai tájkép folyamatos összetettsége csökkenti a termelékenységet és gátolja az innovációt az építőiparban.

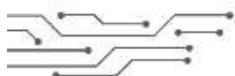
Összefoglalva ezt a részt, érdemes kiemelni a főbb gyakorlati lépéseket, amelyek segítenek a tárgyalt megközelítések alkalmazásában a napi feladatok során:

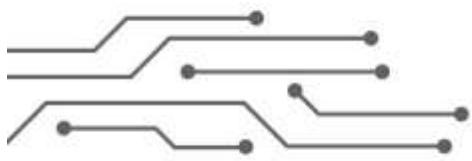
- Az információs tájkép vizualizálása
 - Készítsen vizuális térképet a rendszeresen használt adatforrásokról (Miro, Figma, Canva).
 - Adja hozzá a munkája során használt rendszereket és alkalmazásokat ehhez a térképhez.
 - Azonosítsa a potenciálisan párhuzamos funkciókat és a redundáns megoldásokat.
 - Azonosítani azokat a kritikus pontokat, ahol a rendszerek közötti átvitel során adatvesztés vagy -rongálódás következhet be.
- Személyre szabott adatkezelési gyakorlatok bevezetése
 - Az alkalmazásokról az adatokra, mint a folyamatok kulcsfontosságú eszközeire való összpontosítás

- Dokumentálja az adatforrásokat és a feldolgozási módszertant az átláthatóság biztosítása érdekében
 - Az adatminőség értékelésére és javítására szolgáló mechanizmusok kidolgozása
 - Törekedjen arra, hogy az adatok egyszer kerüljenek bevitelre és többször felhasználásra - ez a hatékony folyamatszervezés alapja.
- Az adatközpontú (adatvezérelt) megközelítés előmozdítása a csapatban
- Javasolja szabványosított és egységes formátumok használatát a peer-to-peer adatcseréhez.
 - Rendszeresen felveti az adatok minőségével és rendelkezésre állásával kapcsolatos kérdéseket a csoportértekezleteken.
 - Ismerje meg a nyílt forráskódú alternatíváit a problémák megoldására használt eszközöknek.

Kezdje kicsiben - válasszon ki egy konkrét folyamatot vagy adathalmazt, amely kritikus a munkája szempontjából, és alkalmazzon adatközpontú megközelítést, és helyezze a hangsúlyt az eszközökről az adatokra. Azzal, hogy sikereket ér el egyetlen kísérleti projektben, nemcsak gyakorlati tapasztalatot szerezhet, hanem egyértelműen demonstrálhatja az új módszertan előnyeit is a csapata számára. A legtöbb lépés elvégzése során, ha kérdései vannak, bármelyik naprakész LLM-től kérhet felvilágosítást és segítséget.

A könyv következő részeiben részletesebben foglalkozunk az adatok strukturálási és harmonizációs technikáival, és a heterogén információk integrálásának gyakorlati megközelítéseit vizsgáljuk. Különös figyelmet szentelünk a széttagolt silókról az egységesített adatökoszisztémákra való áttérésnek, amelyek kulcsszerepet játszanak az építőipar digitális átalakulásában.





III RÉSZ

ADATKERET AZ ÉPÍTŐIPARI ÜZLETI FOLYAMATOKBAN

A harmadik részben az építési adatok tipológiájának és hatékony szervezésük módszereinek átfogó megértése alakul ki. A strukturált, strukturálthatlan, félig strukturált, szöveges és geometriai adatokkal való munka jellemzőit és sajátosságait elemizzük az építési projektekkel összefüggésben. Áttekintésre kerülnek az ágazatban használt különböző rendszerek közötti információcserére szolgáló modern tárolási formátumok és protokollok. A több formátumú adatok egyetlen strukturált környezetbe történő átalakításának gyakorlati eszközeit és technikáit ismertetjük, beleértve a CAD (BIM) adatok integrálásának módját. Javaslatot tesznek a adatok minőségének szabványosításán és validáláson keresztül történő biztosítására, ami kritikus fontosságú az építési számítások pontossága szempontjából. Részletesen elemzik a modern technológiák (Python Pandas, LLM -modellek) használatának gyakorlati szempontjait kódpéldákkal az építőipar tipikus problémáinak megoldására. Megalapozott a kompetenciaközpont (CoE) mint az információkezelési megközelítések összehangolását és szabványosítását szolgáló szervezeti struktúra létrehozásának értéke.

FEJEZET 3.1.

ADATTÍPUSOK AZ ÉPÍTŐIPARBAN

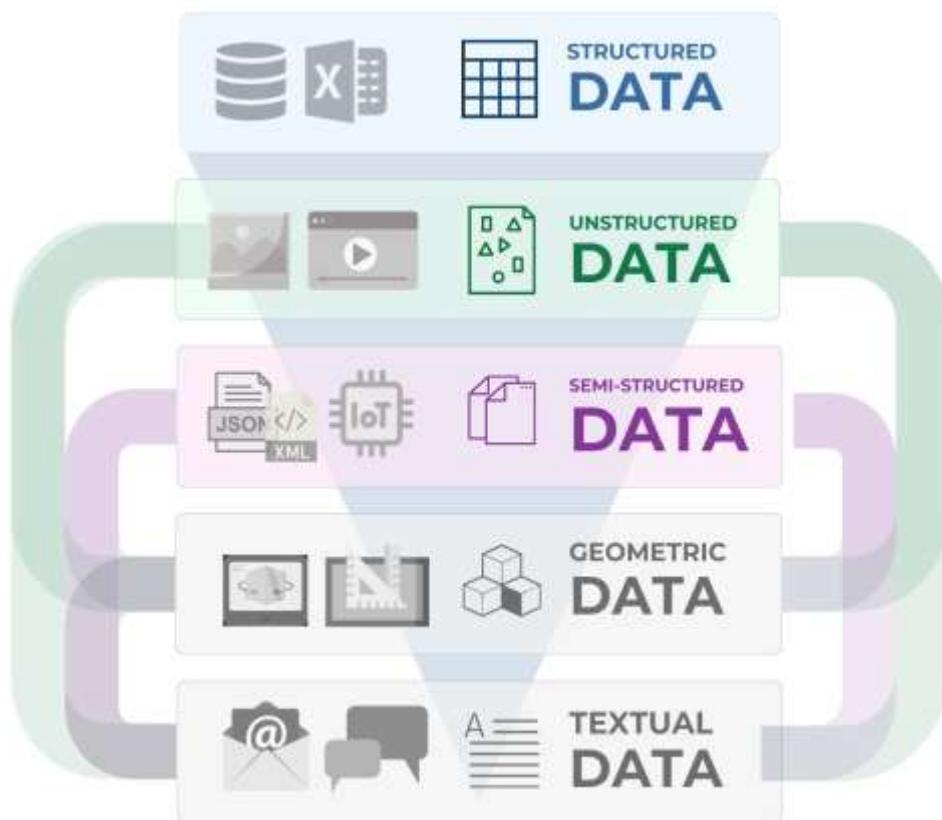
A legfontosabb adattípusok az építőiparban

A modern építőiparban a cégek rendszerei, alkalmazásai és adattárai aktívan töltődnek fel különböző típusú és formátumú információkkal és adatokkal (3.1-1. ábra). Nézzük meg közelebbről a főbb adattípusokat, amelyek egy modern, az építőiparban működő vállalat információs tájképét alkotják:

- **Strukturált** adatok: ezek az adatok világos szervezeti struktúrával rendelkeznek, pl. Excel táblázatok és relációs adatbázisok.
- **Strukturálatlan** adatok: ezek olyan információk, amelyek nincsenek szigorú szabályok szerint rendszerezve. Ilyen adat például a szöveg, a videó, a fényképek és a hangfelvételek.
- **Lazán strukturált** adatok: ezek az adatok a strukturált és a strukturálatlan adatok között helyezkednek el. Tartalmazzák a struktúra elemeit, de ez a struktúra nem minden egyértelmű, vagy gyakran különböző sémákkal írják le. A főleg strukturált adatok példái az építőiparban: műszaki leírások, projektdokumentáció vagy előrehaladási jelentések.
- **Szöveges adatok:** Ide tartozik minden, ami szóbeli és írásbeli kommunikációból származik, például e-mailek, értekezletek és találkozók jegyzőkönyvei.
- **Geometriai** adatok: ezek az adatok a CAD programokból származnak, amelyekben a szakemberek a projektelemek geometriai adatait hozzák létre a megjelenítés, a térfogatértékek megerősítése vagy az ütközések ellenőrzése céljából.

Fontos megjegyezni, hogy a geometriai és szöveges (alfanumerikus) adatok nem külön kategóriát alkotnak, hanem minden adattípusban előfordulhatnak. A geometriai adatok például minden strukturált adatok (parametrikus CAD formátumok), minden strukturálatlan adatok (szkennelt rajzok) részét képezhetik. A szöveges adatok hasonlóképpen lehetnek adatbázisokba rendezettek (strukturált adatok) és létezhetnek egyértelmű struktúra nélküli dokumentumokként is.

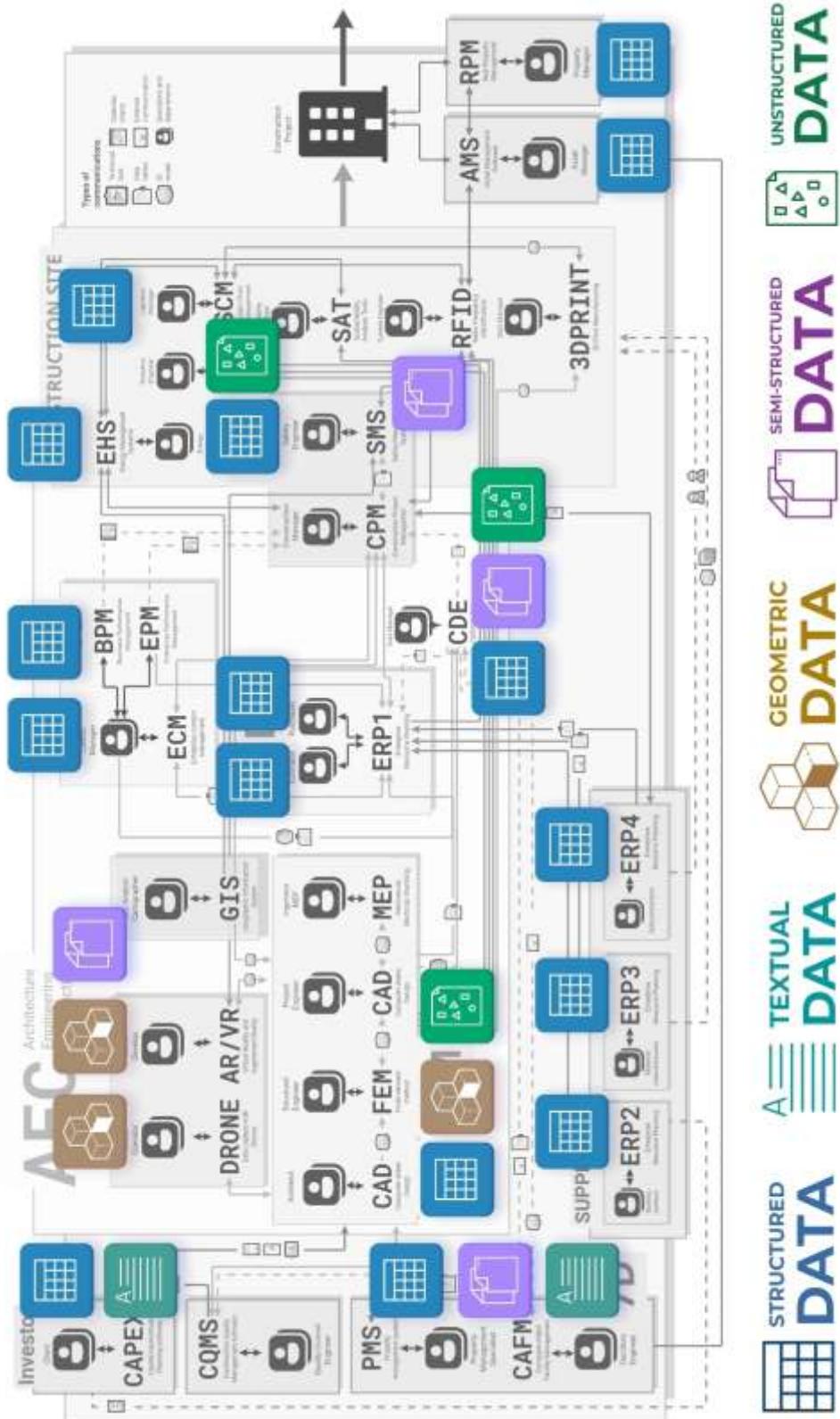
Egy építőipari vállalatnál minden egyes adattípus egyedi elemet képez a vállalat információs eszközeinek mozaikjában. A strukturálatlan adatoktól kezdve, mint például az építkezésekben készült képek és a megbeszélések hangfelvételei, a strukturált nyilvántartásokig, beleértve a táblázatokat és adatbázisokat, minden egyes elem fontos szerepet játszik a vállalat információs tájképének kialakításában.



Ábra: 3.1-1 A mérnököknek és az adatkezelőknek meg kell tanulniuk az építőiparban használt valamennyi adattípussal dolgozni.

Az alábbiakban az építőiparban használt rendszerek és a hozzájuk tartozó adattípusok közül csak néhányat mutatunk be (3.1-2. ábra):

- **ERP** (Enterprise Resource Planning) - általában strukturált adatokat kezel a vállalati erőforrások kezelése és a különböző üzleti folyamatok integrálása érdekében.
- **CAD** (Computer-Aided Design) kombinálva **BIM-mel** (Building Information Modeling) - geometriai és félig strukturált adatokat használ az építési projektek tervezéséhez és modellezéséhez, biztosítva az információk pontosságát és konziszenciáját a tervezési fázisban.
- **GIS** (Geographic Information Systems) - geometriai és strukturált adatokkal dolgozik, hogy kartográfiai adatokat és térbeli kapcsolatokat hozzon létre és elemezzen.
- **RFID** (rádiófrekvenciás azonosítás) - félig strukturált adatokat használ az építkezésen lévő anyagok és berendezések hatékony nyomon követésére rádiófrekvenciás azonosítással.
- **Az ECM** (Engineering Content Management) a mérnöki adatok és dokumentáció kezelésére szolgáló rendszer, beleértve a félig strukturált és strukturálatlan adatokat, például a műszaki rajzokat és tervdokumentációkat.

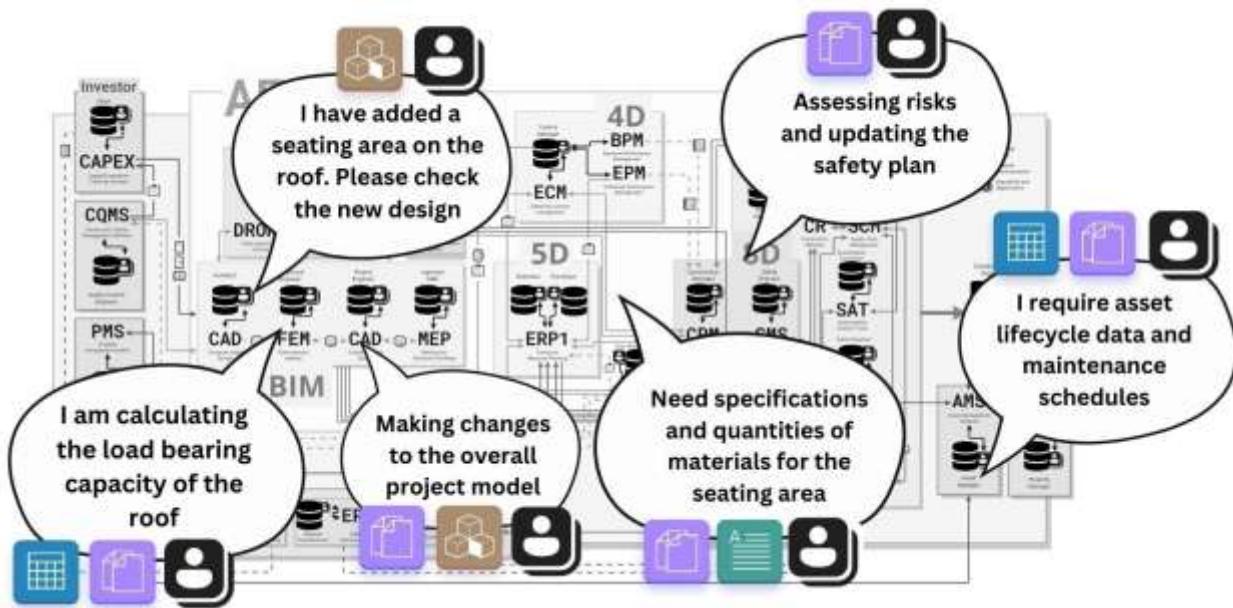


Ábra: 3.1-2 Különböző formátumok és adatok népesítik be a különböző rendszereket, amelyek komplex integrációra alkalmas formába való fordítást igényelnek.

Ezek és a vállalat számos más rendszere az adatok széles skáláját kezeli, a strukturált táblázatos adatoktól az összetett geometriai modellekig, integrált együttműködést biztosítva a tervezési, tervezési és építésirányítási folyamatokban.

Az egyszerűsített párbeszéd példájában (3.1-3. ábra) különböző típusú adatok cseréjére kerül sor az építési projekt szakemberei között:

- ⦿ **Építész:** "Az ügyfél kívánságait figyelembe véve a tetőn egy ülőhelyet alakítottam ki. Kértem, tekintse meg az új tervet" (geometriai adatok - modell).
- ⦿ **Szerkezeti mérnök:** "A projektet megkaptuk. Számítom a tető teherbírását az új üdülöterület számára" (strukturált és félig strukturált adatok - számítási táblázatok).
- ⦿ **Beszerzési vezető:** "Szükségem van a rekreációs területre vonatkozó anyagok specifikációira és mennyiségeire a beszerzés megszervezéséhez" (szöveges és félig strukturált adatok - listák és specifikációk).
- ⦿ **Egészségügyi és biztonsági mérnök:** "Megkapta az új területre vonatkozó adatokat. Felmérem a kockázatokat és frissítem a biztonsági tervet" (félig strukturált adatok - dokumentumok és tervezek).
- ⦿ **A BIM szakértője** -modellezés: "A teljes projektmodell módosításainak elvégzése a munkadokumentáció kiigazítása érdekében" (geometriai adatok és félig strukturált adatok).
- ⦿ **Projektmenedzser:** "Az új pihenőhelyet beépítem a munkaprogramba. Frissítem az ütemterveket és az erőforrásokat a projektirányítási rendszerben" (strukturált és félig strukturált adatok - ütemtervezek és tervezek).
- ⦿ **Létesítménykarbantartási szakember (FM):** "Előkészítem az adatokat a rekreációs terület jövőbeli karbantartásához, és beviszem azokat a vagyonkezelő rendszerbe" (strukturált és félig strukturált adatok - utasítások és karbantartási tervezek).



3.1-3. ábra A szakemberek közötti kommunikáció szöveg- és adatszinten egyaránt zajlik.

Mindegyik szakember más-más típusú adatokkal dolgozik a hatékony csapatmunka és a projekt sikerének biztosítása érdekében. A strukturált, félig strukturált és strukturálatlan adatok közötti különbségek megértése lehetővé teszi, hogy felismerje az egyes típusok egyedi szerepét a digitális üzleti folyamatokban. Nemcsak azt fontos tudni, hogy különböző adatformák léteznek, hanem azt is, hogy hogyan, hol és miért

használják őket.

Nem is olyan régen még az ilyen sokféle adat kombinálásának ötlete ambiciózusnak, de nehezen megvalósíthatónak tűnt. Ma már a minden napjai gyakorlat része. A különböző sémájú és szerkezetű adatok integrálása a modern információs rendszerek architektúrájának szerves részévé vált.

A következő fejezetekben részletesen megvizsgáljuk azokat a kulcsfontosságú szabványokat és megközelítéseket, amelyek lehetővé teszik a strukturált, félig strukturált és strukturálatlan adatok egyetlen összefüggő nézettérrel történő egyesítését. Különös figyelmet fordítunk a strukturált adatokra és a relációs adatbázisokra, mint az építőiparban az információk tárolásának, feldolgozásának és elemzésének fő mechanizmusairól.

Strukturált adatok

Az építőiparban az információ számos forrásból származik - rajzokból, műszaki leírásokból, ütemtervekből és jelentések ből. Ahhoz, hogy ezt az információáramlást hatékonyan lehessen kezelní, strukturálni kell. A strukturált adatok lehetővé teszik az információk kényelmes, olvasható és hozzáférhető formában történő rendszerezését.

A JB Knowledge 5. éves építőipari technológiai jelentése [17] szerint az építőipari projektmenedzsment szakemberek 67%-a manuálisan vagy táblázatok segítségével követi és értékeli a munkateljesítményt.

A legelterjedtebb strukturált adatformátumok közé tartozik az XLSX és a CSV. Ezeket széles körben használják a táblázatokban lévő információk tárolására, feldolgozására és elemzésére. Az ilyen táblázatokban az adatok sorok és oszlopok formájában jelennek meg, ami megkönnyíti az olvasást, szerkesztést és elemzést.

XLSX, a Microsoft által létrehozott **formátum** az XML-struktúrák használatán alapul, és a ZIP algoritmus segítségével archiválódik. A formátum fő jellemzői:

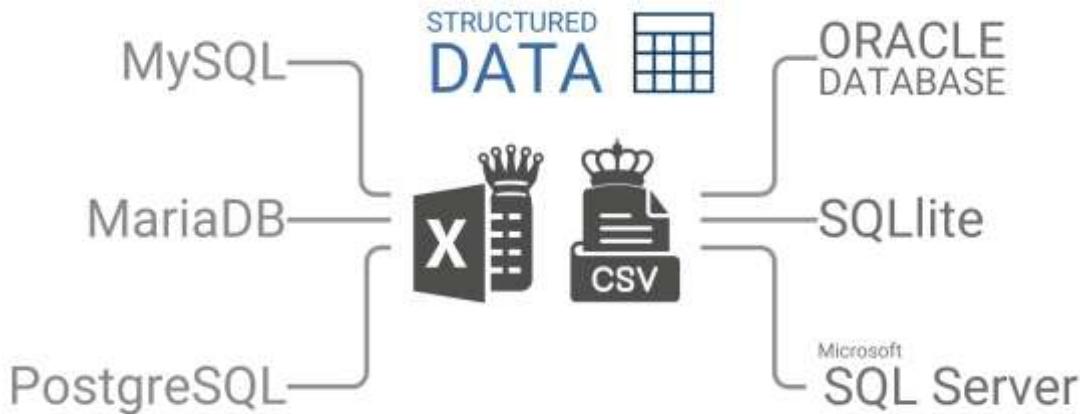
- Komplex képletek, diagramok és makrók támogatása.
- Az adatok különböző lapokon való tárolásának képessége, valamint az információk formázása.
- Optimalizálva a Microsoft Excel, de kompatibilis más irodai programcsomagokkal is.

CSV formátum egy egyszerű szöveges fájl, amelyben az értékek vesszővel, pontosvesszővel vagy más elválasztó karakterekkel vannak elválasztva. Főbb előnyei:

- Univerzális kompatibilitás a különböző programokkal és operációs rendszerekkel.
- Egyszerű importálás/exportálás adatbázisokba és analitikai rendszerekbe.
- Könnyű feldolgozás akár szövegszerkesztőkben is.

A CSV azonban nem támogatja a képleteket és a formázást, így fő alkalmazási területe a rendszerek közötti adatcseré és az információk tömeges frissítése. Sokoldalúsága és platformfüggetlensége miatt a CSV a heterogén informatikai környezetekben történő adatátvitel népszerű eszközévé vált.

Az XLSX és a CSV két formátum összekötő kapocsként szolgál a strukturált adatokkal foglalkozó különböző rendszerek között (3.1-4. ábra). Különösen hasznosak azokban a feladatokban, ahol fontos az olvashatóság, a kézi szerkesztés és az alapvető kompatibilitás.



Ábra 3.1-4 Az XLSX és a CSV formátum a strukturált adatokkal dolgozó különböző rendszerek közötti összekötő kapocs.

A platformfüggetlenség teszi a CSV a legnépszerűbb formátumot a heterogén informatikai környezetekben és rendszerekben történő adatátvitelhez.

Az XLSX és a CSV azonban nem nagy teljesítményű számításokhoz vagy nagy mennyiségű adat hosszú távú tárolásához készült. Ilyen célokra modernebb strukturált formátumokat használnak, mint például az Apache Parquet, Apache ORC, Feather, HDF5. Ezeket a formátumokat részletesebben a könyv 9. részének "Nagy adatok tárolása: népszerű formátumok és hatékonyságuk elemzése" című fejezetében tárgyaljuk.

A gyakorlatban az Excel XLSX formátumú gyakrabban használják kisebb feladatokra és rutinfolyamatok automatizálására. Az összetettebb forgatókönyvek olyan adatkezelési rendszerek használatát igénylik, mint az ERP, PMIS CAFM, CPM, SCM és mások (3.2-1. ábra). Ezek a rendszerek strukturált adatokat tárolnak, amelyeken a vállalat információáramlásának szervezése és irányítása alapul.

Az építőiparban használt modern adatkezelési információs rendszerek táblázatokba rendezett, strukturált adatokra épülnek. A nagy mennyiségű információ megbízható, skálázható és holisztikus kezelése érdekében az alkalmazás- és rendszerfejlesztők a relációs adatbázis-kezelő rendszerek (RDBMS) felé fordulnak.

Relációs adatbázisok RDBMS és az SQL lekérdezési nyelv

Az adatok hatékony tárolása, feldolgozása és elemzése érdekében a **relációs adatbázisok (RDBMS)** olyan adattároló rendszerek, amelyek az információkat táblákba rendezik, amelyek között meghatározott

kapcsolatok vannak.

Az adatbázisokban (RDBMS) tárolt adatok nem csupán digitális információk, hanem a tranzakciók és a különböző rendszerek közötti kölcsönhatások alapját is képezik.

Íme néhány a legelterjedtebb relációs adatbázis-kezelő rendszerek (RDBMS) közül (3.1-5. ábra):

- **MySQL** (nyílt forráskódú) A az egyik legnépszerűbb RDBMS, amely a LAMP stack (Linux, Apache, MySQL, PHP /Perl/Python) része. Egyszerűsége és nagy teljesítménye miatt széles körben használják a webfejlesztésben.
- **PostgreSQL** (nyílt forráskódú) egy nagy teljesítményű objektum-relációs rendszer, amely megbízhatóságáról és fejlett funkcióiról ismert. Alkalmas komplex vállalati megoldásokhoz.
- **A Microsoft SQL Server** a Microsoft kereskedelmi rendszere, amelyet a vállalati környezetekben széles körben használnak a vállalat más termékeivel való integrálhatósága és magas szintű biztonsága miatt.
- **Az Oracle Database** az egyik legerősebb és legmegbízhatóbb DBMS, amelyet nagyvállalatok és kritikus fontosságú alkalmazások használnak.
- **IBM DB2** - nagyvállalatoknak szánt, nagy teljesítményt és hibatűrést biztosító rendszer.
- **Az SQLite** (nyílt forráskódú) egy könnyű beágyazott adatbázis, amely ideális mobil alkalmazásokhoz és önálló rendszerekhez, például CAD-tervezési szoftverekhez (BIM).

Az építőiparban népszerű adatbázis-kezelő rendszerek - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 és SQLite - strukturált adatokkal dolgoznak. Mindezek az DBMS-ek nagy teljesítményű és rugalmas megoldások az üzleti folyamatok és alkalmazások széles körének kezelésére, a kis honlapoktól a nagyvállalati rendszerekig (3.2-1. ábra).

A Statista [48] szerint a relációs adatbázis-kezelő rendszerek (RDBMS) a 2022-ben használatban lévő összes DBMS mintegy 72%-át teszik ki.

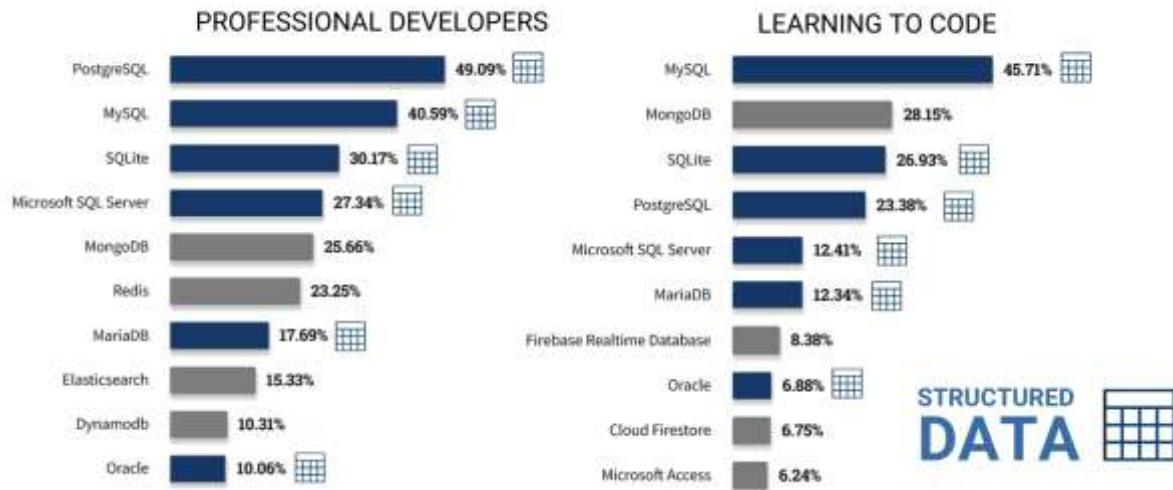
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open Source
	5.	5.	5.	MongoDB	Document, Multi-model	Open Source
	6.	7.	9.	Snowflake®	Relational	Commercial
	7.	6.	6.	Redis®	Key-value, Multi-model	Open Source
	8.	8.	7.	Elasticsearch®	Multi-model	Open Source
	9.	9.	8.	IBM Db2	Relational, Multi-model	Commercial
	10.	10.	10.	SQLite	Relational	Open Source
	11.	11.	12.	Apache Cassandra®	Multi-model	Open Source
	12.	12.	11.	Microsoft Access®	Relational	Open Source
	13.	13.	17.	Databricks®	Multi-model	Commercial
	14.	14.	13.	MariaDB	Relational, Multi-model	Open Source
	15.	15.	14.	Splunk	Search engine	Commercial
	16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
	17.	17.	15.	Microsoft Azure SQL	Relational, Multi-model	Commercial

3.1-5. ábra A strukturált adatbázisok (kékkel jelölve) használatának népszerűsége a DBMS-ek rangsorolásában ([49] alapján).

A nyílt forráskódú adatbázisok telepítése - akár széleskörű technikai ismeretek nélkül is - igen egyszerű. A nyílt forráskódú rendszerek, például a PostgreSQL, a MySQL vagy az SQLite, ingyenesen elérhetőek, és a legtöbb operációs rendszeren működnek: Windows, macOS és Linux. Mindössze annyit kell tennie, hogy felkeresi a projekt hivatalos weboldalát, letölți a telepítőt, és követi az utasításokat. A legtöbb esetben a telepítés nem tart tovább 10-15 percnél. Egy ilyen adatbázist fogunk modellezni és létrehozni a könyv negyedik részében (4.3-8. ábra).

Ha cége felhőszolgáltatásokat használ (például Amazon Web Services, Google Cloud vagy Microsoft Azure), az adatbázist néhány kattintással telepítheti - a platform kész sablonokat kínál a telepítéshez. A kód nyitottságának köszönhetően az ilyen adatbázisok könnyen testre szabhatók az Ön feladataihoz, és a hatalmas felhasználói közösség mindig segít megoldást találni bármilyen problémára.

Az RDBMS továbbra is számos olyan üzleti alkalmazás és elemzési platform alapja (3.1-6. ábra), amelyek lehetővé teszik a vállalatok számára az adatok hatékony tárolását, feldolgozását és elemzését - és ezáltal a megalapozott és időszerű döntések meghozatalát.



3.1-6 A StackOverflow (a legnagyobb informatikai fórum) fejlesztői körében végzett felmérés arról, hogy mely adatbázisokat használtak tavaly, és melyeket akár ják használni jövőre (az RDBMS-ek kékkel vannak kiemelve) ([50] alapján).

Az RDBMS megbízhatóságot, adatkonzisztenciát és tranzakciótámgatást biztosít, és nagy teljesítményű lekérdezési nyelvet - SQL (Structured Query Language) - használ, amelyet gyakran használnak az analitikában, és amely lehetővé teszi az adatbázisokban tárolt információk egyszerű megszerzését, módosítását és elemzését. Az SQL a relációs rendszerekben az adatokkal való munka fő eszköze.

SQL - lekérdezések az adatbázisokban és új trendek

A relációs adatbázisokban gyakran használt SQL nyelv fő előnye más típusú információkezeléssel szemben (például a klasszikus Excel táblázatok segítségével) a nagyon nagy mennyiségű adatbázisok nagy sebességű lekérdezés-feldolgozásának támogatása.

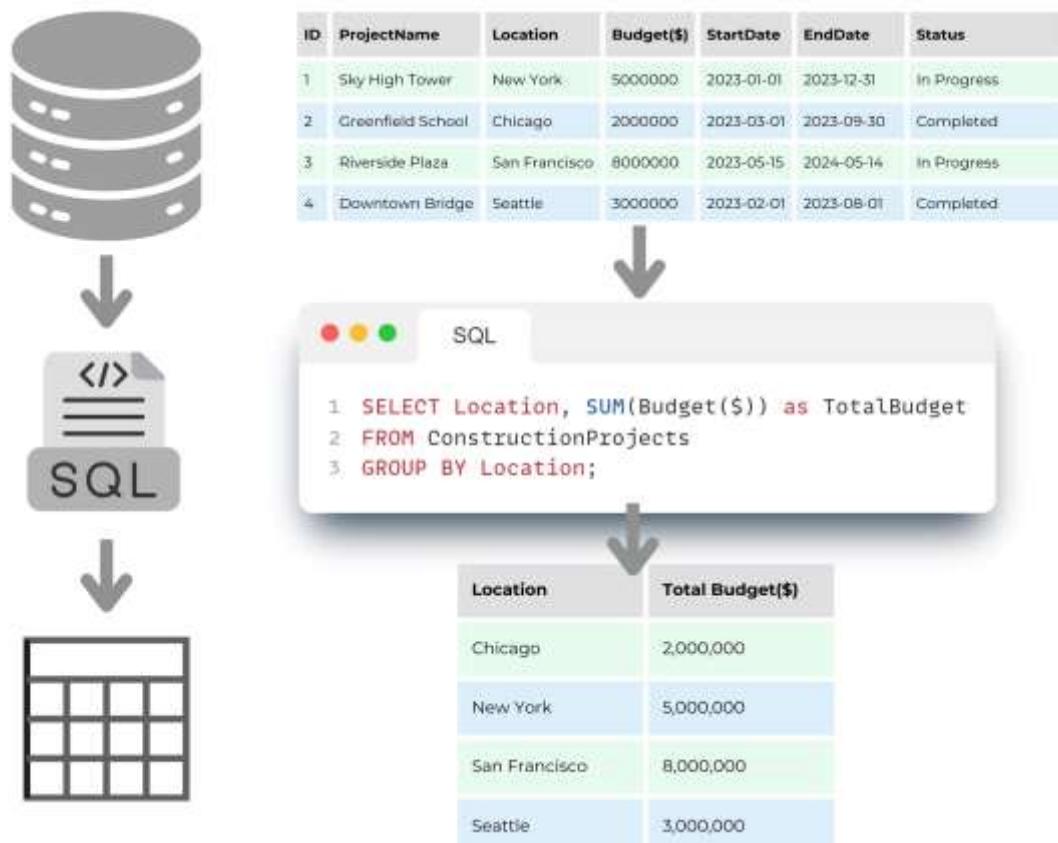
A strukturált lekérdezési nyelv (SQL) egy speciális programozási nyelv, amelyet a relációs adatbázisokban lévő információk tárolására, feldolgozására és elemzésére terveztek. Az SQL az adatok létrehozására, kezelésére és elérésére szolgál, lehetővé téve az információk hatékony keresését, szűrését, kombinálását és összesítését. Az adatok elérésének kulcsfontosságú eszközöként szolgál, kényelmes és formalizált módot biztosít az információraktárakkal való interakcióra.

A SEQUEL-SQL rendszerek fejlődése olyan jelentős termékeken és vállalatokon keresztül halad, mint az Oracle, az IBM DB2, a Microsoft SQL Server, az SAP, a PostgreSQL és a MySQL, és az SQLite és a MariaDB [51] megjelenésében csúcsosodik ki. Az SQL olyan táblázatkezelési képességeket biztosít, amelyek nem találhatók meg az Excelben, így az adatmanipuláció skálázhatóbbá, biztonságosabbá és könnyen automatizálhatóvá válik.

- **Adatszerkezetek létrehozása és kezelése (DDL):** Az SQL oldalon létrehozhat, módosíthat és törlhet táblákat egy adatbázisban, létrehozhatja a táblák közötti kapcsolatokat és meghatározhatja az adattárolási struktúrákat. Az Excel ezzel szemben rögzített lapokkal és

cellákkal dolgozik, anélkül, hogy a lapok és az adathalmazok között egyértelműen meghatározott kapcsolatokat hozna létre.

- **Adatmanipuláció (DML):** Az SQL lehetővé teszi az adatok nagy sebességgel történő tömeges hozzáadását, módosítását, törlését és lekérdezését, valamint összetett lekérdezések végrehajtását szűréssel, rendezéssel és táblázatok összekapcsolásával (3.1-7. ábra). Az Excelben nagy mennyiségű információ feldolgozása kézi műveleteket vagy speciális makrókat igényel, ami lelassítja a folyamatot és növeli a hibák valószínűségét.
- **Hozzáférés-szabályozás (DCL):** Az SQL lehetővé teszi az adatokhoz való hozzáférési jogok differenciálását a különböző felhasználók számára, korlátozva az információk szerkesztésének vagy megtekintésének lehetőségét. Az Excelben ezzel szemben a hozzáférés vagy megosztott (a fájl átvitelekor), vagy összetett beállításokat igényel a jogosultságok megosztásával a felhőszolgáltatásokon keresztül.



Ábra 3.1-7 Példa DML-re az SQL-ben: gyors feldolgozás, csoportosítás és aggregálás néhány sornyi kóddal az automatikus adatfeldolgozáshoz.

Az Excel vizuális és intuitív felépítése miatt megkönnyíti az adatokkal való munkát. Az adatmennyiségek növekedésével azonban az Excel teljesítménye csökken. Az Excel a tárolható adatmennyiségre vonatkozó korlátokkal is szembesül - legfeljebb egymillió sor -, és a teljesítmény már jóval e határ előtt romlik. Tehát míg az Excel előnyösebbnek tűnik a kis adatmennyiségek megjelenítéséhez és kezeléséhez, az SQL jobban alkalmas a nagy adathalmazok kezelésére.

A strukturált adatok fejlődésének következő állomása az oszlopos adatbázisok (Columnar Databases)

megjelenése volt, amelyek a hagyományos relációs adatbázisok alternatíváját jelentik, különösen, ha lényegesen nagyobb adatmennyiségekről és analitikus számításokról van szó. A soros adatbázisokkal ellentétben, ahol az adatokat soronként tárolják, az oszlopos adatbázisok oszloponként rögzítik az információkat. A klasszikus adatbázisokhoz képest ez lehetővé teszi:

- Csökkentse a tárhelyet az oszlopok egységes adatainak hatékony tömörítésével.
- Felgyorsítja az analitikus lekérdezéseket, mivel csak a szükséges oszlopok kerülnek beolvasásra, nem pedig a teljes táblázat.
- A Big Data és az adattárházak optimalizálása, pl. Data Lakehouse Architecture.

Az oszlopos adatbázisokról, a Pandas DataFrame, az Apache Parquet, a HDF5, valamint az ezeken alapuló Big Data -tárolók adatalemzési és -feldolgozási célú létrehozásáról bővebben e könyv következő fejezeteiben - "DataFrame: egy univerzális táblázatos adatformátum" és "Adattárolási formátumok és munka az Apache Parquet: DWH -adattárházak és Data Lakehouse architektúra " - fogunk beszélni.

Strukturálatlan adatok

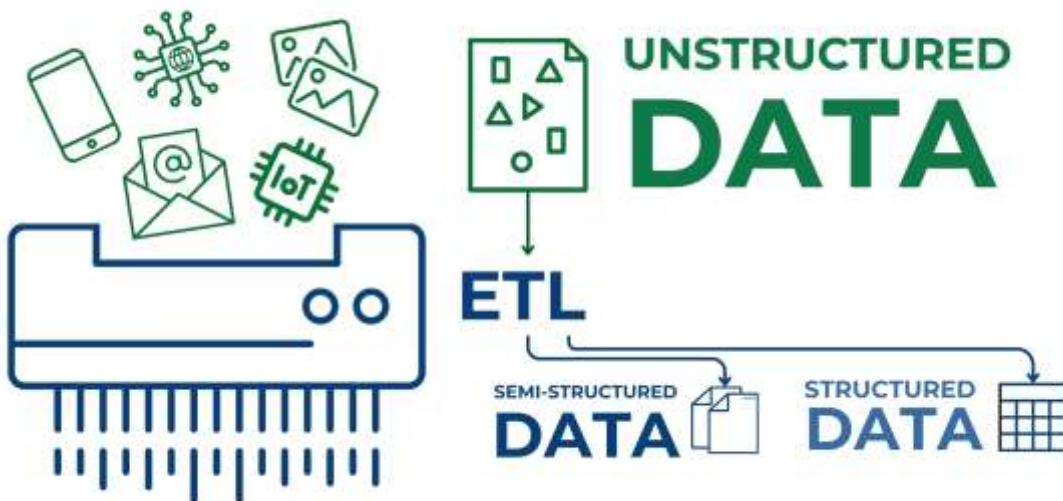
Bár az alkalmazásokban és információs rendszerekben használt adatok többsége strukturált formában van, az építőiparban keletkező információk többsége strukturálatlan adat - képek, videók, szöveges dokumentumok, hangfelvételek és egyéb tartalmi formák - formájában keletkezik. Ez különösen igaz az építési, üzemeltetési és műszaki felügyeleti szakaszokra, ahol a vizuális és szöveges információk dominálnak.

A strukturálatlan adatok olyan információk, amelyeknek nincs előre meghatározott modellje vagy struktúrája, és nem a hagyományos sorok és oszlopok szerint vannak elrendezve, mint az adatbázisokban vagy táblázatokban.

Általánosságban a strukturálatlan adatok két kategóriába sorolhatók:

- Ember által generált strukturálatlan adatok, amelyek magukban foglalják az ember által generált különböző típusú tartalmakat: szöveges dokumentumok, e-mailek, képek, videók stb.
- A gépek által generált strukturálatlan adatokat az eszközök és érzékelők hozzák létre: ezek közé tartoznak például a naplófájlok, a GPS-adatok, a tárgyak internetének eredményei (IoT) és más telemetriai információk egy építkezésről.

A strukturált adatokkal ellentétben, amelyek kényelmesen táblázatokba és adatbázisokba rendezhetők, a strukturálatlan adatok további feldolgozási lépésekkel igényelnek az információs rendszerekbe való integrálásuk előtt (3.1-8. ábra). Az ilyen adatok automatizált gyűjtésére, elemzésére és átalakítására szolgáló technológiák alkalmazása új lehetőségeket nyit az építési hatékonyság javítására, a hibák csökkentésére és az emberi tényező hatásának minimalizálására.



3.1-8 ábra A strukturálatlan adatok feldolgozása a félig strukturált és strukturált adatokkal való átalakítással kezdődik.

A strukturálatlan adatok teszik ki a vállalatok szakemberei által megismert információk akár 80%-át [52], ezért a könyv következő fejezeteiben részletesen, példákkal illusztrálva tárgyaljuk típusukat és feldolgozásukat.

A könnyebb érthetőség kedvéért a szöveges adatokat külön kategorizáljuk. Bár ezek inkább strukturálatlan adatok típusát képezik, fontosságuk és elterjedtségük az építőiparban külön figyelmet igényel.

Szöveges adatok: a strukturálatlan káosz és a strukturált káosz között y

A szöveges adatok az építőiparban az információk formátumának és típusának széles skáláját foglalják magukban, a papíralapú dokumentumuktól az informális kommunikációs módszerekig, mint például a levelek, beszélgetések, munkahelyi levelezés és szóbeli megbeszélések az építkezésen. Mindezek a szöveges adatok fontos információkat hordoznak az építési projektek irányítása szempontjából, a tervezési döntések és a tervmódosítások részleteitől a biztonsági kérdések megvitatásáig és a vállalkozókkal és ügyfelekkel folytatott tárgyalásokig (3.1-9. ábra).



3.1-9. ábra Szöveges adatok, a projekt résztvevői közötti kommunikációban használt információk egyik legnépszerűbb típusa.

A szöveges információk lehetnek formalizáltak és strukturálatlanok egyaránt. A formalizált adatok közé tartoznak a Word-dokumentumok (.doc,.docx), a PDF-ek, valamint az ülések jegyzőkönyveit tartalmazó szöveges fájlok (.txt). A nem formalizált adatok közé tartoznak a messenger és a levelezés, az értekezletek átiratai (Teams, Zoom, Google Meet), valamint a megbeszélések hangfelvételéi (.mp3,.wav), amelyeket szöveggé kell alakítani.

Míg azonban az írásos dokumentumok, például a hivatalos kérelmek, a szerződési feltételek és az e-mailek általában már rendelkeznek bizonyos struktúrával, a szóbeli kommunikáció és a munkaköri levelezés gyakran strukturálatlan marad, ami megnehezíti elemzésüket és integrálásukat a projektmenedzsment-rendszerekbe.

A szöveges adatok hatékony kezelésének kulcsa az adatok strukturált formátumba történő átalakítása. Ez lehetővé teszi a feldolgozott információk automatikus integrálását a meglévő rendszerekbe, amelyek már strukturált adatokkal dolgoznak.



3.1-10. ábra Szöveges tartalom strukturált adattá alakítása.

A szöveges információk hatékony felhasználásához azokat automatikusan strukturált formába kell alakítani (3.1-10. ábra). Ez a folyamat általában több lépésből áll:

- **Szövegfelismerés (OCR)** - dokumentumok és rajzok képének gépi olvasásra alkalmas formátumba történő átalakítása.
- **Szövegelemzés (NLP)** - a kulcsfontosságú paraméterek (a projekt szempontjából releváns dátumok, összegek és számok) automatikus azonosítása.
- **Adatok osztályozása** - az információk kategorizálása (pényüzgy, logisztika, kockázatkezelés).

A felismerés és osztályozás után a már strukturált adatok adatbázisokba integrálhatók, és automatizált jelentési és irányítási rendszerekben használhatók.

Félig strukturált és lazán strukturált adatok

Félig strukturált adatok tartalmaznak bizonyos szintű szervezést, de nem rendelkeznek szigorú sémával vagy struktúrával. Bár az ilyen információk strukturált elemeket tartalmaznak (pl. dátumok, alkalmazottak nevei és az elvégzett feladatok listái), a megjelenítési forma projektről projektre vagy akár alkalmazottról alkalmazottra is jelentősen eltérhet. Ilyen adatok például az időnaplók, az előrehaladási jelentések és az ütemtervek, amelyek különböző formátumokban jeleníthetők meg.

A félig strukturált adatok könnyebben elemezhetőek, mint a strukturálatlan adatok, de további feldolgozást igényelnek a szabványosított projektirányítási rendszerekbe való integráláshoz.

A félig strukturált adatokkal való munka, amelyeket folyamatosan változó struktúra jellemz, jelentős kihívásokat jelent. Ennek oka, hogy az adatszerkezet változékonysága miatt a félig strukturált adatok egyes forrásainak feldolgozásához és elemzéséhez külön-külön megközelítésekre van szükség.

Míg azonban a strukturálatlan adatok kezelése sok erőfeszítést igényel, a félig strukturált adatok feldolgozása viszonylag egyszerű módszerekkel és eszközökkel elvégezhető.

Gyengén strukturált adatok egy általánosabb kifejezés, amely a minimális vagy hiányos struktúrájú adatokat írja le. Ezek leggyakrabban szöveges dokumentumok, csevegések, e-mailek, amelyekben néhány metaadat (pl. dátum, feladó) megtalálható, de az információ nagy része kaotikusan van megjelenítve.

Az építőiparban a lazán strukturált adatok számos folyamatban előfordulnak. Ezek közé tartozhatnak például:

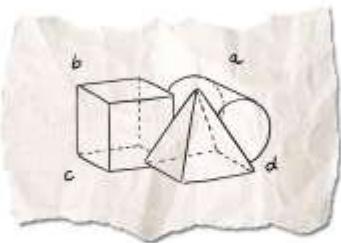
- Becslések és árajánlatok - táblázatok az anyagokra, mennyiségekre és költségekre vonatkozó adatokkal, de egységes formátum nélkül.
- Rajzok és műszaki diagramok - PDF vagy DWG formátumú fájlok, amelyek szöveges megjegyzéseket és metaadatokat tartalmaznak, de nem rendelkeznek szigorúan rögzített struktúrával.
- Munkatervez - adatok az MS Project, Primavera P6 vagy más rendszerekből, amelyek eltérő exportstruktúrával rendelkezhetnek.
- CAD (BIM -modellek) - tartalmazzák a szerkezet elemeit, de az adatok megjelenítése a szoftvertől és a projekt szabványától függ.

A CAD-rendszerek által előállított geometriai adatok ugyanúgy kategorizálhatók, mint a félig strukturált adatok. A geometriai CAD (BIM) adatokat azonban külön adattípusként fogjuk besorolni, mivel a szöveges adatokhoz hasonlóan gyakran külön adattípusként kezelhetők a vállalati folyamatokban.

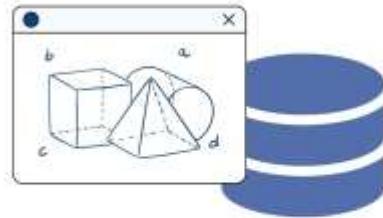
Geometriai adatok és alkalmazása

Míg a projektelemekre vonatkozó metaadatokat szinte minden táblázatok, strukturált vagy lazán strukturált formátumok formájában tárolják, a projektelemek geometriai adatait a oldalon a legtöbb esetben speciális CAD eszközökkel hozzák létre (3.1-11. ábra), amelyek lehetővé teszik a projektelemek részletes megjelenítését vonalak (2D) vagy geometriai testek (3D) halmazaként.

3000 BCE - 1960s

physical medium
(artefact)

1960s to present day

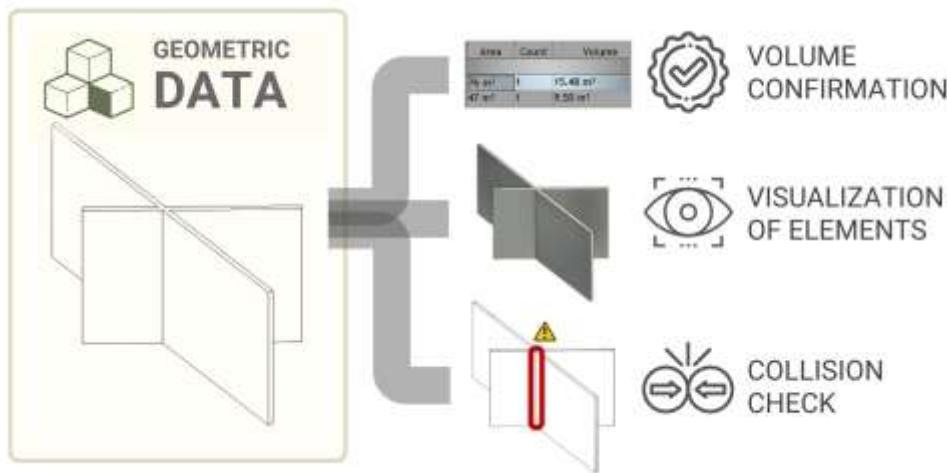
digital project data
(CAD data)

3.1-11. ábra CAD eszközök segítettek a geometriai információk fizikai adathordozóról adatbázisba történő átvitelében.

Az építőiparban és az építészetben a geometriai adatokkal való munka során a geometriai adatok három fő felhasználási területe különböztethető meg (3.1-12. ábra):

- **A térfogatok megerősítése:** a projektelemek térfogatának és méreteinek automatikus és pontos meghatározásához geometriai adatokra van szükség, amelyeket a CAD programokban (BIM) speciális geometriai magok segítségével generálnak. Ezek az adatok tartalmazzák az automatikusan kiszámított területeket, térfogatokat, hosszokat és egyéb fontos attribútumokat, amelyek a tervezéshez, a költségvetés összeállításához és az erőforrások és anyagok megrendeléséhez szükségesek
- **A projekt vizualizációja:** a projektben bekövetkező bármilyen változás esetén az elemek vizualizációja lehetővé teszi a különböző síkokban frissített rajzok automatikus generálását. A projekt kezdeti szakaszokban történő vizualizációja lehetővé teszi a gyorsabb megértést az összes résztvevő között, így időt és erőforrásokat takaríthat meg az építési folyamat során.
- **Ütközések ellenőrzése:** Az összetett építési és mérnöki projekteknél, ahol több elemkategória (pl. csövek és falak) "geometriai konfliktusok" nélküli kölcsönhatása kritikus fontosságú, az ütközések ellenőrzése kulcsszerepet játszik. A ütközésellenőrző szoftver használatával proaktívan azonosíthatja a projektelemek közötti potenciális geometriai konfliktusokat, megelőzve ezzel a költséges hibákat az építési folyamat során.

A mérnöki tervezőirodák kezdeteitől, az első komplex szerkezetek építésének idejétől kezdve a statikusok geometriai információkat szolgáltattak rajzok, vonalak és sík geometriai elemek formájában (papiruszon, "A0" papíron vagy DWG, PDF, PLT formátumban), amelyek alapján a művezetők és a becslők (3.1-11. ábra), az elmúlt évezredekre vonalzók és szögmérők segítségével gyűjtötték az attribútumokat vagy az elemek és elemcsoportok számát. 3.1-11), az elmúlt évezredekre során vonalzó és szállítók segítségével attributív térfogatokat vagy elemek és elemcsoportok mennyiségették gyűjtötték.



Ábra 3.1-12 A geometria az alapja az elemek térfogatparamétereinek meghatározásához, amelyeket aztán a projekt költségeinek és ütemezésének kiszámításához használnak.

Manapság ezt a kézi és időigényes feladatot teljes automatizálással oldják meg, köszönhetően a modern CAD-eszközökben megjelenő volumetrikus modellezésnek (BIM), amely egy speciális geometriai mag segítségével automatikusan lehetővé teszi bármely elem volumetrikus attribútumainak meghatározását anélkül, hogy a volumetrikus paramétereket kézzel kellene kiszámítani.

A modern CAD-eszközök lehetővé teszik a projektelemek osztályozását és kategorizálását is, így a projektadatbázisból specifikációs táblázatok tölthetők fel különböző rendszerekbe, például költségbecsleshez, ütemezéshez vagy CO₍₂₎-számításhoz (3.1-13. ábra). A specifikációk, a mennyiségi táblázatok és a mennyiségek beszerzését és gyakorlati példákat a "Mennyiségek beszerzése és mennyiségi számítás" című fejezetben tárgyaljuk.



Ábra 3.1-13 A CAD-eszközök (BIM) az adatokat olyan adatbázisokban tárolják, amelyeket más rendszerekkel való integrálásra és kölcsönhatásra terveztek.

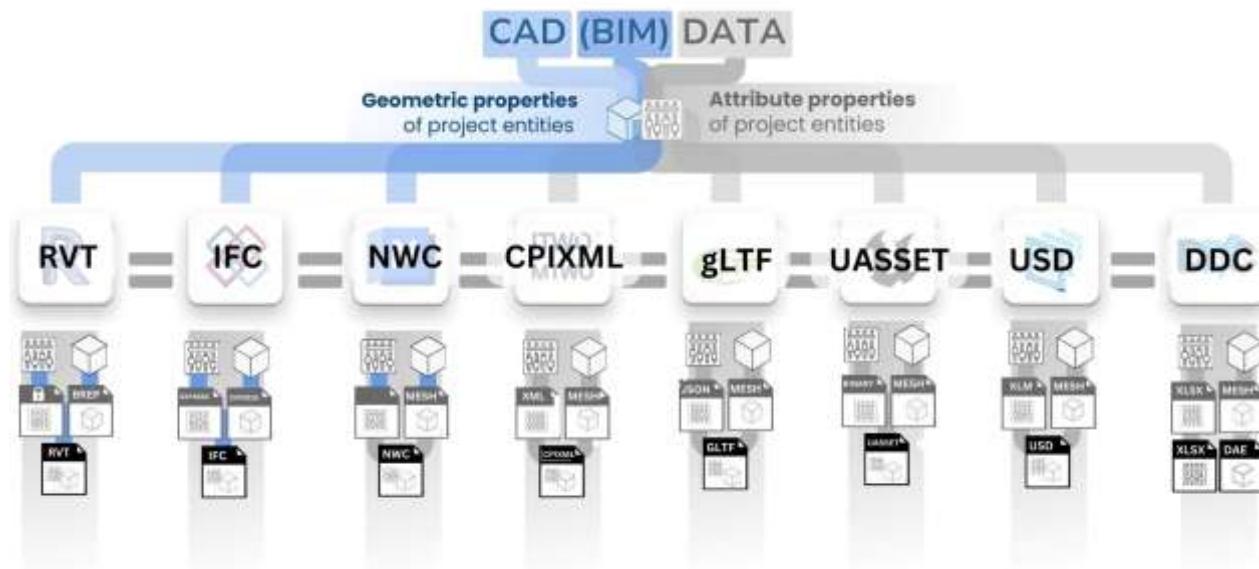
A CAD-környezetben használt adatbázisok és formátumok zárt jellege miatt a CAD-megoldásokban létrehozott geometriai adatok valójában különálló információtípussá váltak. Az elemek geometriáját és a (strukturált vagy félig strukturált) metainformációkat egyaránt magában foglalja, amelyeket speciális fájlokba és formátumokba zárnak.

CAD adatok: a tervezéstől az adattárolásig

A modern CAD és BIM rendszerek saját, gyakran szabadalmaztatott formátumokban tárolják az adatokat: DWG, DXF, RVT, DGN, PLN és mások. Ezek a formátumok támogatják az objektumok 2D-s és 3D-s ábrázolását egyaránt, és nemcsak a geometriát, hanem az objektumokhoz kapcsolódó attribútumokat is megőrzik. Itt vannak a leggyakoribbak:

- **A DWG** egy bináris fájlformátum, amelyet kétdimenziós (és ritkábban háromdimenziós) tervezési adatok és metaadatok tárolására használnak.
- **A DXF** egy szöveges formátum a 2D és 3D -rajzok CAD -rendszerek közötti cseréjére. Geometriát, rétegeket és attribútumadatokat tartalmaz, támogatja az ASCII és a bináris ábrázolást is.
- **Az RVT** egy bináris formátum a CAD modellek tárolására, beleértve a 3D -geometriát, az elemek attribútumait, a kapcsolatokat és a tervezési paramétereket.
- **Az IFC** egy nyílt szöveges formátum az építési adatok CAD (BIM) rendszerek közötti cseréjére. Tartalmazza a geometriát, az objektumtulajdonságokat és a köztük lévő kapcsolatokra vonatkozó információkat.

Ezeken kívül más formátumokat is használnak: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Bár céljuk és nyitottsági szintük eltérő (3.1-14. ábra), mind ugyanazt a projektinformációs modellt különböző formában tudják megjeleníteni. Összetett projektekben ezeket a formátumokat gyakran párhuzamosan használják, a rajzolástól a projektmodellek koordinálásáig.



3.1-14. ábra A népszerű CAD-tárolási formátumok a geometriát BREP- vagy MESH-paraméterekkel írják le, amelyeket attribútumadatok egészítenek ki.

A fenti formátumok mindegyike lehetővé teszi az építési projekt minden egyes elemére vonatkozó adatok tárolását, és a fenti formátumok mindegyike két kulcsfontosságú adattípust tartalmaz:

- **Geometriai paraméterek** - egy objektum alakját, helyét és méreteit írják le. A geometriát és annak használatát részletesen a könyv CAD (BIM) megoldásoknak szentelt hatodik részében tárgyaljuk;
- **Attribútumtulajdonságok** - különböző információkat tartalmaznak: anyagok, elemtípusok, műszaki

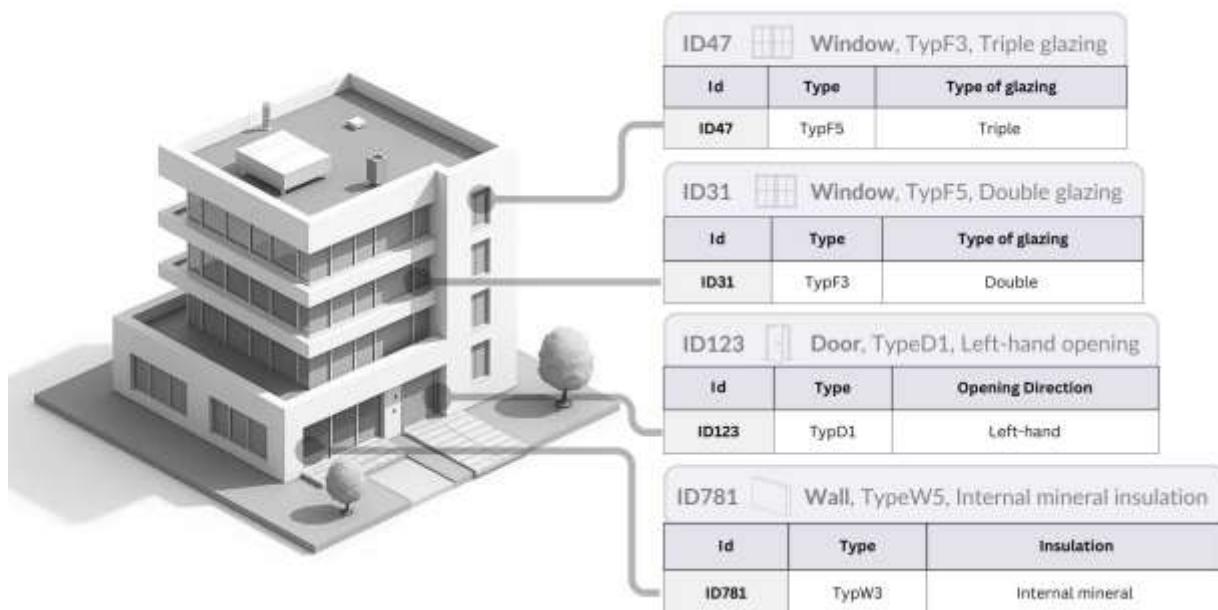
jellemzők, egyedi azonosítók és egyéb tulajdonságok, amelyekkel a projektelemek rendelkezhetnek.

Az attribútumadatok különösen fontosak a modern projektekben, mivel meghatározzák az objektumok működési jellemzőit, lehetővé teszik a mérnöki és költségszámításokat, valamint a tervezés, az építés és az üzemeltetés résztvevői közötti végponttól végpontig tartó interakciót. Például:

- Ablakok és ajtók esetében: szerkezeti típus, üvegezés típusa, nyitási irány (3.2-1. ábra).
- A falak esetében az anyagokra, a hőszigetelésre és az akusztikai teljesítményre vonatkozó információkat rögzítik.
- A mérnöki rendszerek esetében a csővezetékek, csatornák, kábelvezetékek és csatlakozásaik paraméterei kerülnek tárolásra.

Ezek a paraméterek tárolhatók mind a CAD-(BIM-)fájlokban, mind külső adatbázisokban - exportálás, konverzió vagy a belső CAD-struktúrákhoz való közvetlen hozzáférés eredményeként, reverse engineering eszközökkel. Ez a megközelítés megkönnyíti a tervezési információk integrálását más vállalati rendszerekkel és platformokkal

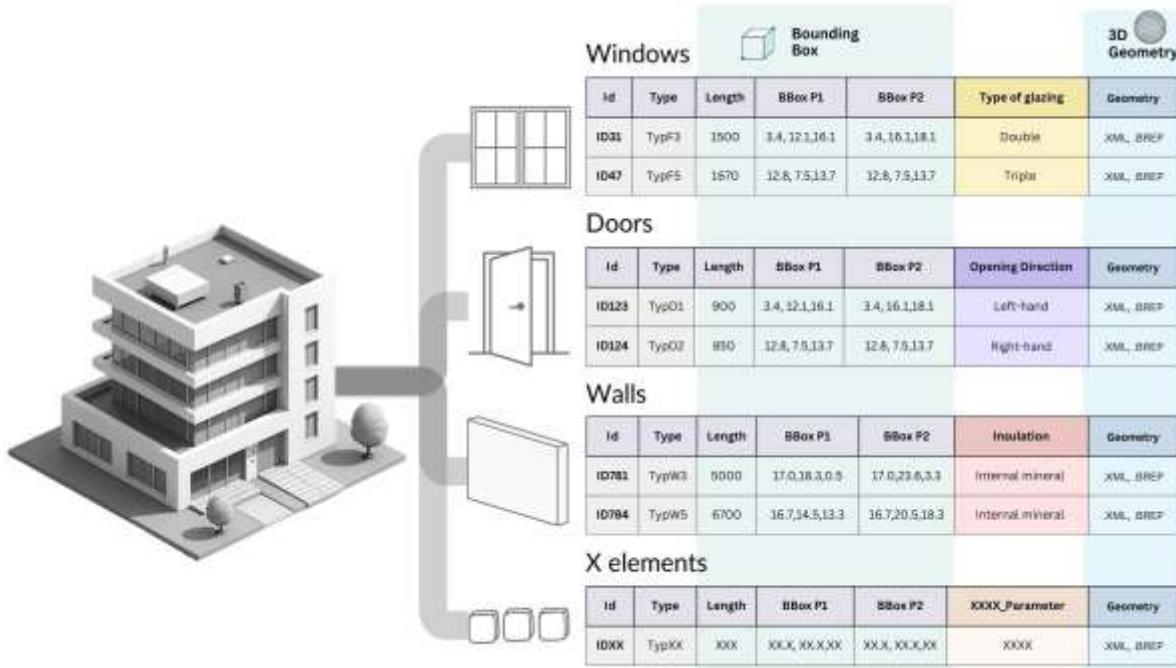
A CAD (BIM) kontextusában **a visszafejtés** egy digitális modell belső szerkezetének kinyerése és elemzése annak érdekében, hogy az eredeti algoritmusokhoz vagy dokumentációhoz való hozzáférés nélkül újra létrehozzák annak logikáját, adatszerkezetét és függőségeit.



Ábra 3.1-15 Egy projektelem a parametrikus vagy poligonális geometria leírásán túlmenően az elemek paramétereiről és tulajdonságairól szóló információkat is tartalmaz.

Ennek eredményeképpen minden egyes elem körül egyedi paraméterek és tulajdonságok alakulnak ki, beleértve az egyes objektumok egyedi jellemzőit (pl. azonosító és méretek) és az elemcsoportok közös attribútumait. Ez lehetővé teszi nemcsak a projekt egyes elemeinek-elemeinek elemzését, hanem azok logikai csoportokba történő összevonását is, amelyeket aztán más szakemberek felhasználhatnak a feladataikhoz és számításaihoz a rendszerekben és adatbázisokban.

Az entitás a valós világ olyan konkrét vagy absztrakt tárgya, amely egyedileg azonosítható, leírható és adat formájában reprezentálható.



Ábra 3.1-16 minden projektelem tartalmaz olyan attribútumokat, amelyeket vagy a tervező ad meg, vagy a CAD programban számol ki.

Az elmúlt évtizedekben az építőipar számos új CAD (BIM) formátumot fejlesztett ki, amelyek egyszerűsítik az adatok létrehozását, tárolását és továbbítását. Ezek a formátumok lehetnek zárt vagy nyílt, táblázatos, parametrikus vagy grafikus formátumok. Sokféleségük és töredezetségük azonban jelentősen megnehezíti az adatkezelést a projekt életciklusának minden szakaszában. Az építőiparban az információcserére használt főbb formátumok összehasonlító táblázata a 3.1-17 ábrán látható (a teljes verzió elérhető a QR-kódval).

Az interoperabilitás és a CAD-adatokhoz való hozzáférés problémáinak megoldására BIM-menedzsereket (BIM) és koordinátorokat vonnak be, akiknek feladata az export ellenőrzése, az adatok minőségének ellenőrzése és a CAD (BIM) adatok részeinek más rendszerekbe történő integrálása.

A formátumok zárt jellege és összetettsége miatt azonban nehéz automatizálni ezt a folyamatot, ami arra kényszeríti a szakembereket, hogy számos műveletet manuálisan végezzenek el, anélkül, hogy teljes körű, in-line adatfeldolgozási folyamatokat (pipeline) tudnának létrehozni.

3.1-17. ábra táblázat, amely összehasonlítja azokat a főbb adatformátumokat, amelyekben a projektelemek információt tárolják [53].

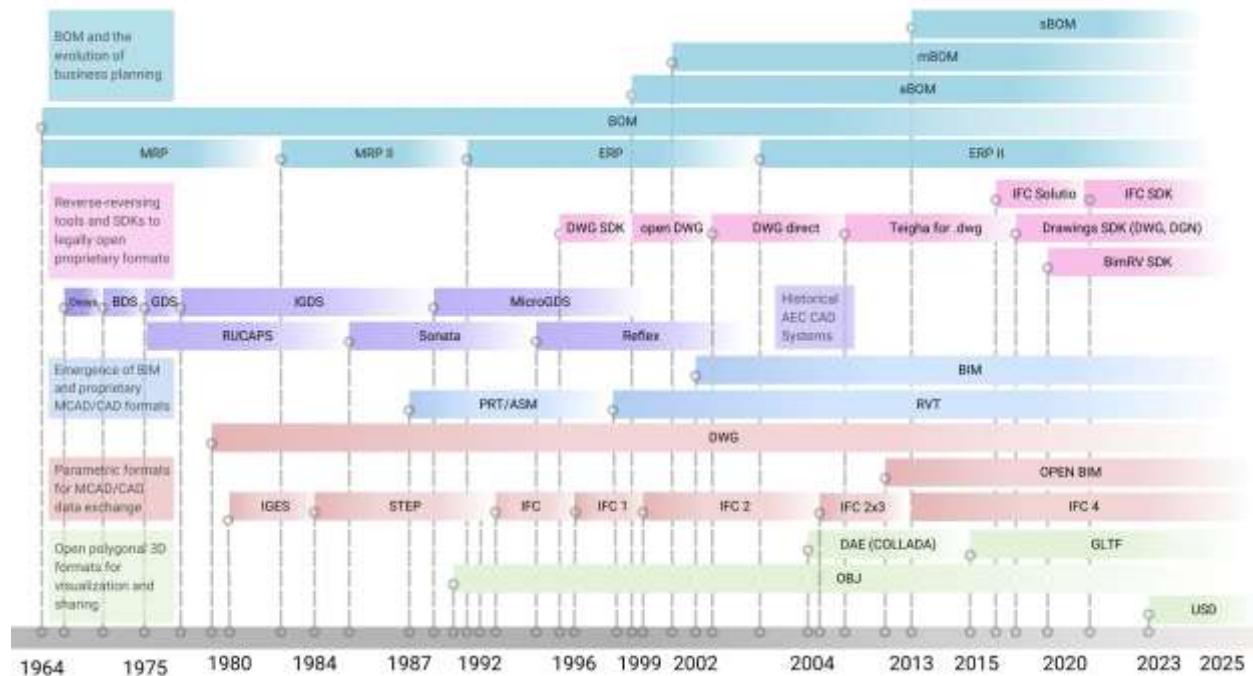
Ahhoz, hogy megértsük, miért van ennyi különböző adatformátum, és miért zárt a legtöbbük, fontos elmagyarázni a CAD (BIM) programokon belül zajló folyamatokban, amelyeket a könyv hatodik részében részletesen megvizsgálunk.

A geometriához hozzáadott további információs réteget a CAD-rendszer fejlesztői vezették be a BIM-fogalom formájában (Building Information Modeling), amely egy 2002 óta aktívan támogatott marketingfogalom az építőiparban [54].

A BIM (BOM) koncepció megjelenése és a CAD felhasználása a folyamatokban.

Az épületinformációs modellezés (BIM) koncepciója, amelyet először a 2002-es BIM Whitepaper [54] vázolt fel, a CAD-szoftverek gyártóinak marketingkezdeményezéseiből származik. A CAD-szoftverek fejlesztőinek marketingkezdeményezéseiből született, és kísérlet volt arra, hogy a gépgézetben már jól bevált elveket az építőipar igényehez igazítsák.

A BIM a BOM (Bill of Materials) - Bill of Materials - koncepciójából származik, amelyet az 1980-as évek vége óta széles körben használnak az iparban. A gépgyártásban a BOM lehetővé tette a CAD-rendszer adatainak összekapcsolását a PDM (Product Data Management), PLM (Product Lifecycle Management) és ERP rendszerekkel, így biztosítva a mérnöki információk holisztikus kezelését a termék teljes életciklusa során (3.1-8. ábra).



Ábra 3.1-18 A műszaki leírások (BOM), az információs modellezés (BIM) és a digitális formátumok fejlődése a mérnöki építőiparban.

A BOM koncepció modern fejlődése egy kibővített keretrendszer - XBOM (Extended BOM) - kialakulásához vezetett, amely nemcsak a termékösszetételeit, hanem viselkedési forgatókönyveket, működési követelményeket, fenntarthatósági paramétereket és a prediktív elemzéshez szükséges adatokat is magában foglalja. Az XBOM lényegében ugyanazt a szerepet tölti be, mint a BIM az építőiparban: minden megközelítés arra törekszik, hogy a digitális modellt a projekt teljes életciklusa során a projekt valamennyi résztvevője számára az igazság egységes forrásává tegye.

A BOM megjelenésének egyik legfontosabb mérföldkőve az építőiparban az első, kifejezetten az építőiparra szabott parametrikus CAD (MCAD) 2002-es bevezetése volt. Ezt az a csapat fejlesztette ki, amely korábban létrehozta a Pro-E®-t, az 1980-as évek végén a gépészeti számára kifejlesztett, forradalmi MCAD-rendszer, amely az iparági szabványnak számított [55].

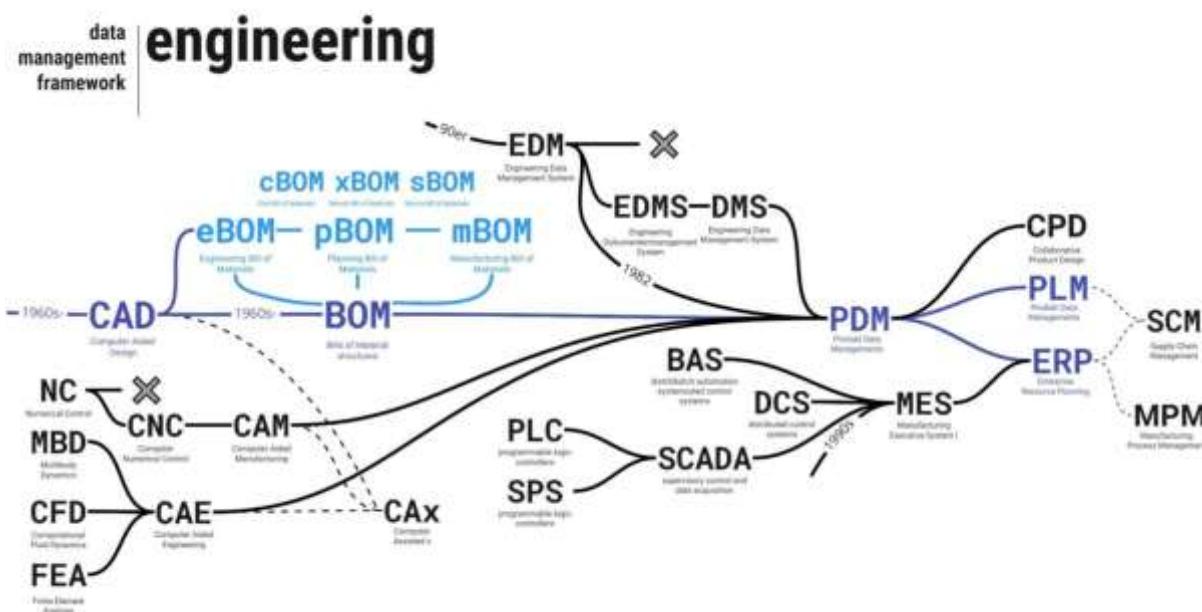
Már az 1980-as évek végén célul tüzték ki az akkori CAD -programok korlátainak [56] kiküszöbölését. A fő cél az volt, hogy csökkentsék a tervezési elemek paramétereinek módosításához szükséges munkaigényt, és lehetővé tegyék a modell frissítését a CAD programokon kívüli adatok alapján, egy adatbázis segítségével [57]. Ebben a legfontosabb szerepet a paraméterezésnek kellett játszania: a jellemzők automatikus lekérdezése az adatbázisból és felhasználása a modell CAD-rendszeren belüli frissítéséhez.

A Pro-E és a mögötte álló elemi parametrikus modellezés c BOM koncepciója jelentős hatással volt a CAD - és az MCAD - piac fejlődésére [58]. Ez a modell 25 éve van jelen az iparban, és számos modern rendszer lett a koncepcionális utódja.

A cél egy olyan rendszer létrehozása, amely elég rugalmas ahhoz, hogy a mérnököket arra ösztönözze, hogy könnyen mérlegeljék a különböző tervezeteket. A terv módosításának költségei pedig a lehető legközelebb legyenek a nullához. A hagyományos CAD / CAM szoftverek irreálisan korlátozzák az olcsó változtatások elvégzését csak a tervezési folyamat legelején [59].

- Samuel Heisenberg, a Parametric Technology Corporation® alapítója, az MCAD -termék Pro-E fejlesztője és az RVT formátumot használó CAD termék megalkotójának tanára.

A gépgyártásban a PDM, PLM, MRP és ERP rendszerek kulcsfontosságú platformokká váltak. Központi szerepet játszanak az adat- és folyamatmenedzsmentben, összegyűjtik a CAx-rendszerből (CAD, CAM, CAE) származó információkat, és a termékstruktúra (BOM: eBOM, pBOM, mBOM) alapján szervezik a tervezési tevékenységeket (3.1-18. ábra). Ez az integráció csökkenti a hibákat, elkerüli az adatok megkettőzését és biztosítja a tervezéstől a gyártásig történő végpontok közötti nyomon követhetőséget.



3.1-19 ábra Történetileg a BOM az 1960-as években jelent meg a CAx-rendszerök adatainak strukturálására és a vezérlőrendszerök felé történő továbbítására.

A korábbi Pro-E csapat által kifejlesztett és a BOM megközelítésén alapuló CAD-megoldás egyik vezető gyártó általi megvásárlását a BIM Whitepaper sorozat (2002-2003)[60][61] szinte azonnali megjelenése jelezte. A BIM koncepciót már a 2000-es évek közepétől kezdve aktívan kezdték népszerűsíteni az építőiparban, ami jelentősen megnövelte a parametrikus szoftverek iránti érdeklődést. A népszerűség olyan gyorsan nőtt, hogy a gépészeti Pro-E építőipari villája - az e gyártó által támogatott parametrikus CAD - ténylegesen kiszorította a versenytársakat az építészeti és statikai tervezés szegmensében (3.1-20. ábra). A 2020-as évek elejére de facto megszilárdította globális dominanciáját a BIM (CAD) piacon [62].



Ábra 3.1-20 Google keresési lekérdezések népszerűsége (RVT versus IFC): a korábbi Pro-E csapat által létrehozott parametrikus CAD BOM támogatással -BIM a világ szinte legtöbb országában népszerűvé vált.

Az elmúlt 20 év során a BIM rövidítés számos értelmezést kapott, amelyek többféle jelentése a 2000-es évek elején kialakult kezdeti marketingkoncepciókban gyökerezik. Az ISO 19650 szabvány, amely fontos szerepet játszott a kifejezés népszerűsítésében, tulajdonképpen biztosította a BIM státuszát, mint az információkezelés "tudományos alapú" megközelítését. Magának a szabványnak a szövegében azonban, amely az objektumok teljes életciklusa során a BIM segítségével történő adatkezeléssel foglalkozik, a BIM rövidítés szerepel ugyan, de soha nem kerül egyértelműen meghatározásra

A gyártó eredeti honlapja, amely 2002-ben[60] és 2003-ban[61] egy sor Whitepaper tett közzé a BIM-ről, valójában a BOM (Bills of Materials) és PLM (Product Lifecycle Management) koncepciókról szóló marketinganyagokat reprodukálta, amelyeket korábban a Pro-E gépészeti szoftverben használtak még az 1990-es években [63].

Az épületinformációs modellezés, az épületek tervezésének, építésének és irányításának innovatív, új megközelítése, amelyet a..... [CAD-gyártó cég neve] vezetett be 2002-ben, világszerte megváltoztatta az iparág szakembereinek gondolkodását arról, hogy a technológia hogyan alkalmazható az épületek tervezésében, építésében és irányításában.

- BIM Whitepaper, 2003 [61].

Ezek a korai publikációk közvetlenül összekapcsolták a BIM-et a központosított integrált adatbázis koncepciójával. A 2003-as Whitepaperben foglaltak szerint a BIM olyan épületinformáció-kezelés, ahol minden frissítés egyetlen tárolóban történik, biztosítva, hogy minden rajz, vágás és specifikáció (BOM - Bills of Materials) szinkronizálva legyen.

A BIM az épületinformáció-kezelés, ahol minden frissítés és minden változás egy adatbázisban történik. Így akár vázlatrajzokról, metszetekről vagy laprajzokról van szó, minden minden összehangolt, következetes és naprakész.

- CAD-cég honlapja vendor with BIM Whitepaper, 2003 [54].

A tervezés egyetlen integrált adatbázis segítségével történő irányításának gondolatát már az 1980-as évek kutatásai során széles körben megvitatták. Charles Eastman BDS koncepciója [57] például 43 utalást tartalmazott az "adatbázis" kifejezésre (6.1-2. ábra). Ez a szám 2004-re csaknem a felére, 23-ra csökkent a BIM-ről szóló 2002-es Whitepaperben [64]. A 2000-es évek közepére pedig az adatbázisok témája gyakorlatilag eltűnt a gyártók marketinganyagaiból és általában a digitalizációs napirendről.

Bár eredetileg az adatbázis és az ahhoz való hozzáférés volt a BIM -rendszer lényege, idővel a hangsúly a geometriára, a vizualizációra és a 3D-re helyeződött át. Maga az IFC szabvány 1994-es regisztrátora, aki 2002-ben a BIM Whitepapert publikálta - ugyanaz a gyártó - a 2000-es évek elején megjelent Whitepaperben kifejezetten rámutatott az IGES, STEP és IFC típusú semleges formátumok korlátaira és a CAD-adatbázisokhoz való közvetlen hozzáférés szükségességére:

A különböző alkalmazások nem kompatibilisek lehetnek, és az újra beírt adatok pontatlannak lehetnek [...]. A hagyományos számítógépes tervezés [CAD] eredménye: magasabb költségek, hosszabb piacra kerülési idő és alacsonyabb termékminőség. Ma már minden fontosabb alkalmazás ipari szabványos interfészeket használ az alacsony szintű adatcseréhez. A régi IGES szabványok vagy az új STEP [az IFC a STEP/IGES formátum de facto és de jure másolata] használatával a különböző gyártók alkalmazásai közötti adatcseréhez a felhasználók elérhetnek bizonyos mértékű adatkompatibilitást a legjobb termékek között. Az IGES és a STEP azonban csak alacsony szinten működik, és nem tudnak olyan gazdag adatot cserélni, mint a mai vezető alkalmazások által generált információk [...]. És bár ezek és más szabványok szinte naponta fejlődnek, az adatgazdagság tekintetében mindig elmaradnak a mai gyártók termékei mögött. [...] Az alkalmazáson belüli programoknak képesnek kell lenniük az adatgazdagság cseréjére és megőrzésére anélkül, hogy semleges fordítókra, például IGES-re, STEP-re [IFC-re] vagy PATRAN-ra szorulnának. Ehelyett a keretalkalmazásoknak képesnek kell lenniük arra, hogy közvetlenül hozzáférjenek a mögöttes CAD-adatbázishoz, hogy az információ részletessége és pontossága ne vesszen el.

- CAD-gyártók Whitepapere (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale," 2000 [65].

Így már az 1980-as években és a 2000-es évek elején a CAD-környezetben a digitális tervezés kulcsfonthosságú elemének az adatbázist tekintették, nem pedig a formátumfájlt vagy a semleges IFC formátumot.

Azt javasolták, hogy a fordítókról le kell mondani, és az alkalmazásoknak közvetlen hozzáférést kell biztosítani az adatokhoz. A valóságban azonban a 2020-as évek közepére a BIM koncepciója az "oszd meg és uralkodj" stratégiához kezdett hasonlítani, ahol a zárt geometriai magokat használó szoftvergyártók érdekeit helyezik előtérbe a nyílt információcsere fejlesztésével szemben.

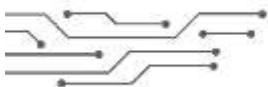
A BIM ma már az építőipar szerves részének tekinthető. Az elmúlt két évtizedben azonban az egyszerűsített együttműködés és adatintegráció igéretei nagyrészt nem váltak valóra. A legtöbb megoldás még mindig zárt vagy semleges formátumokhoz és speciális eszközökhez kötött. A "CAD és BIM: Marketing, valóság és a tervezési adatok jövője az építőiparban" című könyv 6. részében részletesen megvizsgáljuk a BIM történetét, a nyílt BIM-et és az IFC-t, valamint az interoperabilitás és a geometriai magok kérdéseit.

Napjainkban az iparágnak azzal a nagy kihívással kell szembenéznie, hogy a CAD (BIM) hagyományos értelmezéséből, mint modellezési eszközből át kell térnie a teljes értékű adatbázisként való használatra. Ez új megközelítést igényel az információval való munkavégzéshez, a zárt ökoszisztemák függőségének elhagyásához és a nyílt megoldások megvalósításához.

A CAD-adatbázisokhoz való hozzáférést lehetővé tevő reverse engineering eszközök fejlődésével, valamint a nyílt forráskódú és LLM technológiák elterjedésével az építőipari felhasználók és fejlesztők egyre inkább eltávolodnak a szoftvergyártók homályos feltételeitől. Ehelyett a hangsúly arra helyeződik át, ami valóban számít: az adatokra (adatbázisokra) és a folyamatokra.

A divatos rövidítések és vizualizációk mögött a szabványos adatkezelési gyakorlatok állnak: tárolás, átvitel és átalakítás - azaz a klasszikus ETL-folyamat (Extract, Transform, Load). Más iparágakhoz hasonlóan az építőipar digitalizálása nemcsak adatcsereszabványokat, hanem a heterogén információk egyértelműen strukturált kezelését is megköveteli.

A CAD-adatokban (BIM) rejlő lehetőségek teljes körű kiaknázása érdekében a vállalatoknak át kell gondolniuk az információkezeléssel kapcsolatos megközelítésüket. Ez elkerülhetetlenül a digitális átalakulás egyik kulcsfontosságú eleméhez vezet - az építőipari szakemberek által naponta használt adatok egységesítéséhez, szabványosításához és értelmes strukturálásához.

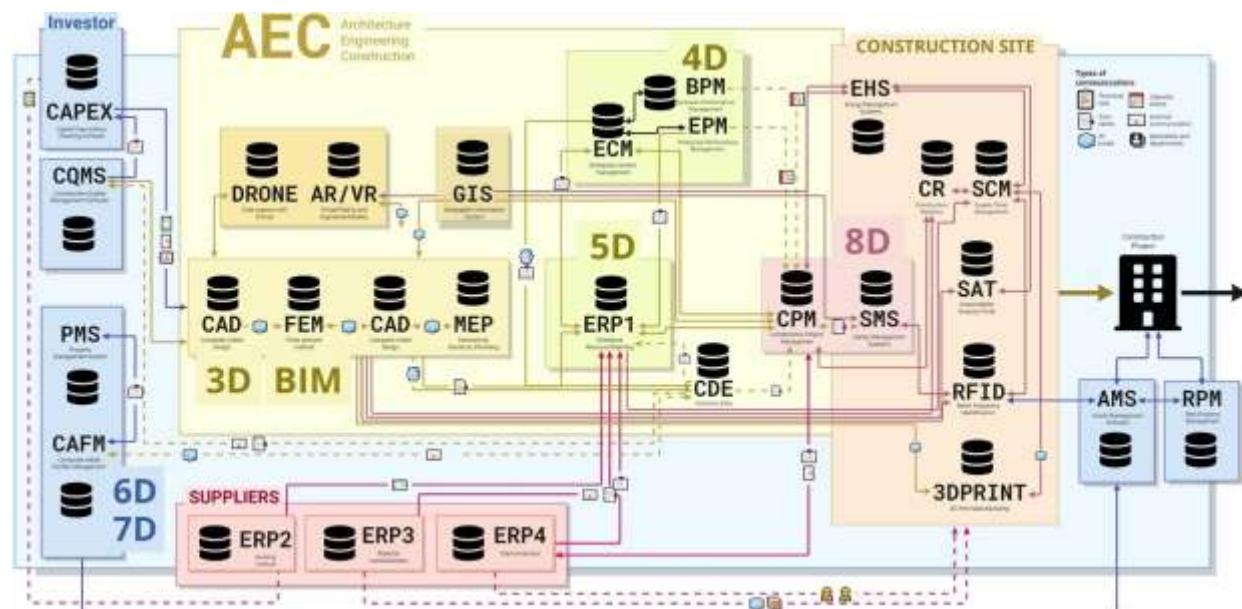


FEJEZET 3.2.

ADATHARMONIZÁCIÓ ÉS STRUKTURÁLÁS

Rendszerek adatokkal való feltöltése az építőiparban

Akár nagyvállalatokról, akár középvállalkozásokról van szó, a szakemberek naponta azzal foglalkoznak, hogy a különböző interfészekkel rendelkező szoftverrendszeret és adatbázisokat több formátumú információkkal tölték meg (3.2-1. ábra), amelyeknek a vezetők segítségével együtt kell működniük egymással. Az egymással kölcsönhatásban lévő rendszerek és folyamatok ezen komplexuma az, amely végső soron bevételt és nyereséget termel a vállalat számára.



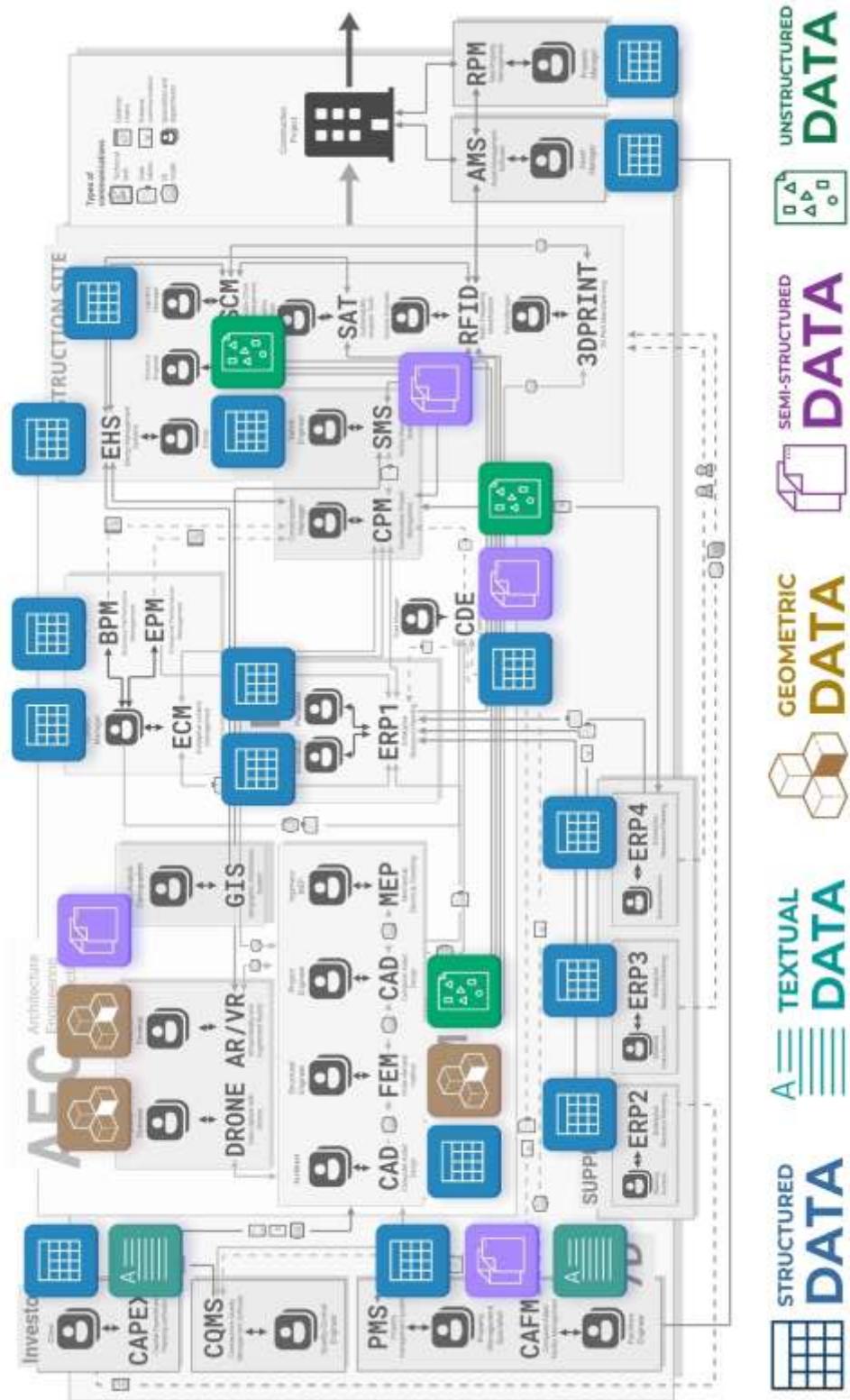
3.2-1 ábra Gyakorlatilag minden, az építőiparban használt rendszer vagy alkalmazás magját a népszerű RDBMS adatbázisok egyike alkotja.

A korábban említett és az építőiparban alkalmazott rendszerek minden egyes kategóriája saját adattípusokkal dolgozik, amelyek megfelelnek e rendszerek funkcionális szerepének. Az absztrakt szintről a konkrét szintre való áttérés érdekében az adattípusoktól a formátumok és dokumentumok formájában történő megjelenítésig jutunk el.

A rendszerek korábban megadott listájához (1.2-4. ábra) most hozzáadjuk az általuk gyakran használt speciális formátumokat és dokumentumtípusokat:

■ Befektető (CAPEX)

- Pénzügyi adatok: költségvetések, kiadási előrejelzések (strukturált adatok).
- Piaci trendadatok: piaci elemzések (strukturált és strukturálatlan adatok).
- Jogi és szerződéses adatok: szerződések (szöveges adatok).



3.2-2. ábra Az építőiparban számos különböző interfésszel rendelkező rendszert használnak, amelyek különböző típusú adatokkal foglalkoznak.

■ Irányítási rendszerek (PMS, CAFM, CQMS)

- Projektadatok: grafikonok, feladatok (strukturált adatok).
- Létesítménykarbantartási adatok: karbantartási tervezek (szöveges és félig strukturált adatok).
- Minőségellenőrzési adatok: szabványok, vizsgálati jelentések (szöveges és strukturálatlan adatok).

■ CAD, FEM és BIM

- Műszaki rajzok: építészeti, szerkezeti tervezek (geometriai adatok, strukturálatlan adatok).
- Modellek építése: 3D -modellek, anyagadatok (geometriai és félig strukturált adatok).
- Mérnöki számítások: terheléselemzés (strukturált adatok).

■ Építkezésirányítási rendszerek (EHS, SCM)

- Biztonsági és egészségügyi adatok: biztonsági protokollok (szöveges és strukturált adatok).
- Ellátási lánc adatok: készletek, megrendelések (strukturált adatok).
- Napi jelentések: munkaidő, termelékenység (strukturált adatok).

■ Drónok, AR/VR, GIS, 3D -nyomtatás

- Geoadatok: topográfiai térképek (geometriai és strukturált adatok).
- Valós idejű adatok: videók és fényképek (strukturálatlan adatok).
- Modellek 3D -nyomtatáshoz: digitális rajzok (geometriai adatok).

■ További irányítási rendszerek (4D BPM, 5D ERP1)

- Idő- és költségadatok: ütemtervezek, becslések (strukturált adatok).
- Változáskezelés: projektváltozások nyilvántartása (szöveges és strukturált adatok).
- Teljesítményjelentés: sikereségi mutatók (strukturált adatok).

■ Adatintegráció és kommunikáció (CDE, RFID, AMS, RPM)

- Adatcsere: dokumentumcsere, adatmodellek (strukturált és szöveges adatok).
- RFID és nyomon követési adatok: logisztika, eszközgazdálkodás (strukturált adatok).
- Monitoring és ellenőrzés: érzékelők a helyszíneken (strukturált és strukturálatlan adatok).

Így az építőiparban minden egyes rendszer - az építésirányítási rendszerektől az operatív adatbázisokig - saját típusú információkkal dolgozik: strukturált, szöveges, geometriai és egyéb információkkal. Az "adattérkép", amellyel a szakembereknek nap mint nap dolgozniuk kell, rendkívül változatos. A formátumok egyszerű felsorolása azonban nem mutatja meg az információkkal való valódi munka összetettségét.

A gyakorlatban a vállalatok azzal szembesülnek, hogy az adatok, még ha a rendszerekből le is hívják őket, nem használhatók fel "úgy, ahogy vannak". Ez különösen igaz a szövegekre, képekre, PDF-ekre, CAD-fájlokra és más olyan formátumokra, amelyeket nehéz a szabványos eszközökkel elemezni. Ezért a következő kulcsfontosságú lépés az adattranszformáció - egy olyan folyamat, amely nélkül a feldolgozás, elemzés, vizualizáció és döntéshozatal nem automatizálható hatékonyan.

Adattranszformáció: a modern üzleti elemzés kritikus alapjai

Napjainkban a legtöbb vállalat paradoxonnal találja magát szemben: napi folyamatainak mintegy 80%-a még mindig a klasszikus strukturált adatokra - a jól ismert Excel-táblázatokra és a relációs adatbázisokra (RDBMS) [66] - támaszkodik. Ugyanakkor azonban a vállalatok digitális ökoszisztémájába belépő új információk 80%-a strukturálatlan vagy lazán strukturált (3.2-3. ábra) [52]. Ide tartoznak a szövegek,

grafikák, geometriai adatok, képek, CAD -modellek, dokumentáció PDF-ben, hang- és videofelvételek, elektronikus levelezés és még sok más.

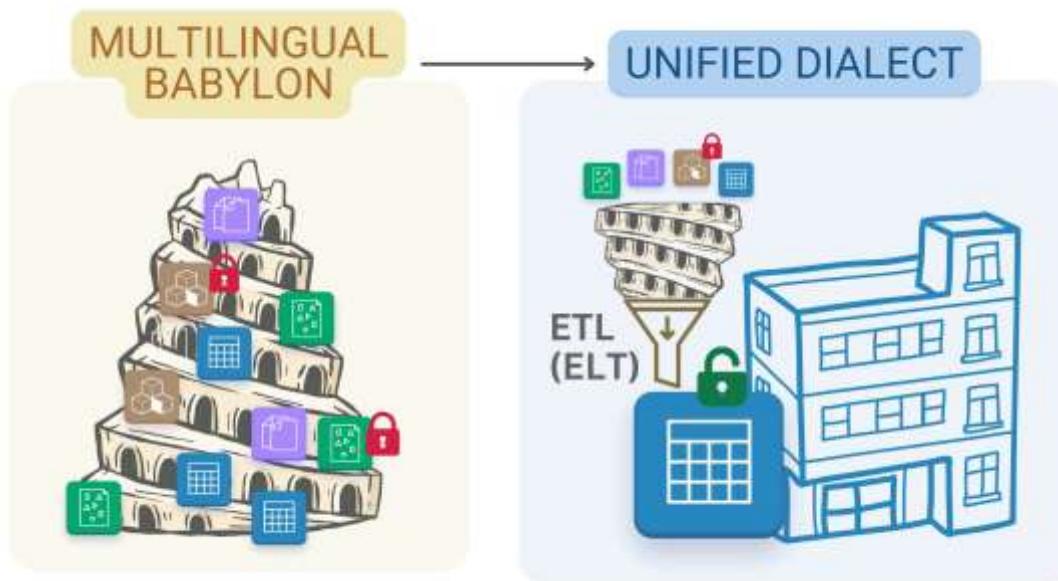
Ráadásul a strukturálatlan adatok mennyisége továbbra is gyorsan növekszik - az éves növekedési ütemet 55-65%-ra becsülik [67]. Ez a dinamika komoly nehézségeket okoz az új információknak a meglévő üzleti folyamatokba való integrálásában. A több formátumú adatok ezen áramlásának figyelmen kívül hagyása információs hiányosságok kialakulásához vezet, és csökkenti a vállalat teljes digitális környezetének kezelhetőségét.



Ábra: 3.2-3 A strukturálatlan adatok éves növekedése kihívást jelent az áramló információk üzleti folyamatokba való integrálásában.

Az összetett, strukturálatlan és zavaros, lazán strukturált adatok figyelmen kívül hagyása az automatizálási folyamatokban jelentős hiányosságokat okozhat a vállalat információs tájképében. A mai kontrollálhatatlan és lavinaszerű információmozgás világában a vállalatoknak olyan hibrid megközelítést kell alkalmazniuk az adatkezelésre, amely hatékony módszereket tartalmaz az összes adattípus kezelésére.

A hatékony adatkezelés kulcsa a szervezésében, strukturálásában és osztályozásában rejlik különböző típusú adatok "Bábel" (beleértve a strukturálatlan, szöveges és geometriai formátumokat, strukturált vagy lazán strukturált adatokká). Ez a folyamat a kaotikus adathalmazokat szervezett struktúrákká alakítja át a rendszerekbe való integráláshoz, és ezáltal lehetővé teszi az ezeken alapuló döntéshozatalt (3.2-4. ábra).



Ábra 3.2-4 Az adatkezelési osztályok fő feladata a sokféle és sokféle formátumú adat "babilonjának" strukturált és kategorizált rendszerré alakítása.

Az ilyen harmonizáció egyik fő akadálya továbbra is a különböző digitális platformok - az előző fejezetekben tárgyalt "silók" - közötti alacsony szintű átjárhatóság.

A jelentés szerint a National Institute of Standards and Technology (NIST, USA) hangsúlyozza [68], hogy a különböző építési platformok közötti rossz adatkompatibilitás információvesztéshez és jelentős többletköltségekhez vezet. Csak 2002-ben a szoftverek interoperabilitási problémái az amerikai tőkeépítésben évente 15,8 milliárd dollár veszteséget okoztak, amelynek kétharmadát az épületek tulajdonosai és üzemeltetői viselik, különösen az üzemeltetés és karbantartás során [68]. A tanulmány azt is megállapítja, hogy az adatformátumok szabványosítása csökkentheti ezeket a veszteségeket és javíthatja a hatékonyságot a létesítmény teljes életciklusa során.

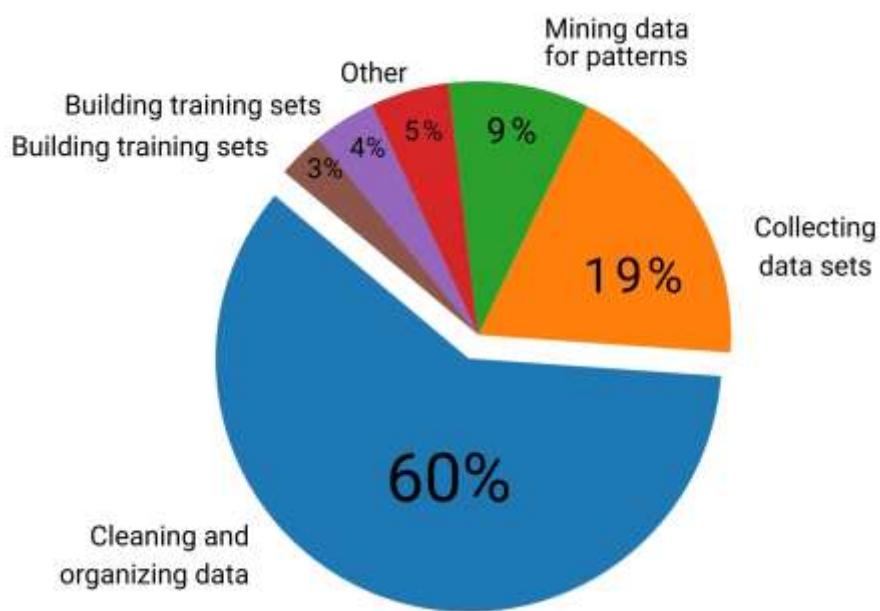
A 2016-os CrowdFlower tanulmány [69] szerint, amely 16 000 adattudósra terjedt ki világszerte, a fő probléma továbbra is a "piszkos" és több formátumú adatok. E tanulmány szerint a legértékesebb erőforrás nem a végleges adatbázisok vagy gépi tanulási modellek, hanem az információk előkészítésére fordított idő.

A tisztítás, formázás és rendszerezés az elemző és az adatkezelő idejének akár 60 százalékát is igénybe veszi. Közel egyötödét a megfelelő adatkészletek keresésével és összegyűjtésével töltik, amelyek gyakran silókban rejtoznak és az elemzéshez hozzáférhetetlenek. És csak az idő körülbelül 9 százalékát fordítják közvetlenül modellezésre, analitikára, előrejelzések készítésére és hipotézisek tesztelésére. A fennmaradó időt a kommunikációval, vizualizálással, jelentéstéssel és a támogató információforrások felkutatásával töltik.

Egy menedzser adatmunkája átlagosan a következőképpen oszik meg (3.2-5. ábra):

- **A adatok tisztítása és rendszerezése (60%):** a tiszta és strukturált adatok birtokában jelentősen csökkenthető a elemzőjének munkaideje, és felgyorsítható a feladatok elvégzése.

- **Adatgyűjtés (19%):** Az adattudományi szakemberek számára az egyik legnagyobb kihívás a releváns adatkészletek megtalálása. A vállalati adatok gyakran kaotikusan szervezett "silókban" vannak egymásra halmozva, ami megnehezíti a szükséges információk elérését.
- **Modellezés/gépi tanulás (9%):** Gyakran hátrálhatja az ügyfelek részéről az üzleti célok tisztázatlansága. A világos küldetésnyilatkozat hiánya még a legjobb modellben rejlő lehetőségeket is semmissé teheti.
- **Egyéb feladatok (5%):** az adatfeldolgozás mellett az elemzőknek kutatással, az adatok különböző szempontok szerinti feltárásával, az eredmények vizualizációkon és jelentéseken keresztül történő közlésével, valamint a folyamatok és stratégiák optimalizálására vonatkozó javaslatokkal kell foglalkozniuk.



Ábra 3.2-5 Mire fordítják a legtöbb időt az adatokkal dolgozó adatkezelők ([70] alapján).

Ezeket a becsléseket más tanulmányok is alátámasztják. A BizReportban 2015-ben közzétett Xplenty tanulmány [71] szerint az üzleti intelligencia (BI) szakemberek (BI) idejének 50-90%-át az adatok elemzésre való előkészítésével töltik.

A adatok tisztítása, validálása és rendszerezése kritikus alapot jelent az összes későbbi adat- és elemzési folyamathoz, és az adattudósok idejének akár 90%-át is igénybe veszi.

Ez a végfelhasználó számára láthatatlan, aprólékos munka kulcsfontosságú. A nyers adatok hibái elkerülhetetlenül torzítják az elemzéseket, félrevezetők és költséges irányítási hibákhoz vezethetnek. Ezért az adattisztítási és szabványosítási folyamatok - a duplikációk kiküszöbölésétől és a hiányosságok pótlásától a mértékegységek harmonizálásáig és a közös modellhez való igazodásig - a modern digitális stratégia sarokkövévé válnak.

Így az adatok alapos átalakítása, tisztítása és szabványosítása nemcsak a szakemberek idejének nagy részét (az adatokkal végzett munka akár 80%-át) foglalja le, hanem meghatározza a modern üzleti folyamatok keretében történő hatékony felhasználás lehetőségét is. Az adatok rendszerezése és tisztítása

önmagában azonban nem meríti ki a vállalat információáramlásának optimális irányításának feladatát. A szervezési és strukturálási szakasz során a megfelelő adatmodell kiválasztása válik szükségessé, amely közvetlenül befolyásolja az információkkal való munka kényelmét és hatékonyságát a feldolgozás későbbi szakaszaiban.

Mivel az adatok és az üzleti célok eltérőek, fontos, hogy megértsük az adatmodellek jellemzőit, és képesek legyünk kiválasztani vagy létrehozni a megfelelő struktúrát. A strukturáltság mértékétől és az elemek közötti kapcsolatok leírásának módjától függően három fő modell létezik: strukturált, lazán strukturált és grafikus. Mindegyik más-más feladatokra alkalmas, és mindegyiknek megvannak a maga erősségei és gyengeségei.

Adatmodellek: kapcsolatok az adatokban és az elemek közötti kapcsolatok

Az információs rendszerekben az adatokat különböző módon szervezik - az információk tárolására, feldolgozására és továbbítására vonatkozó feladatktól és követelményektől függően. Az adatmodellek típusai, az információ tárolásának formája közötti legfontosabb különbség a strukturáltság mértéke és az elemek közötti kapcsolatok leírásának módja.

A strukturált adatok egyértelmű és megismételhető sémával rendelkeznek: táblázatos formában, rögzített oszlopokkal vannak megszervezve. Ez a formátum kiszámíthatóságot, könnyű feldolgozhatóságot és hatékonyságot biztosít az SQL lekérdezések, szűrés és aggregálás során. Példák - adatbázisok (RDBMS), Excel, CSV.

A lazán strukturált adatok rugalmas struktúrát tesznek lehetővé: a különböző elemek különböző attribútumokat tartalmazhatnak, és hierarchiaként tárolhatók. Ilyen például a JSON, az XML vagy más dokumentumformátumok. Ezek az adatok kényelmesek, ha egymásba ágyazott objektumokat és a köztük lévő kapcsolatokat kell modellezni, másrészt azonban bonyolítják az adatelemzést és a szabványosítást (3.2-6. ábra).

	Data Model	Storage Format	Example
	Relational	CSV, SQL	A table of doors in Excel
	Hierarchical	JSON, XML	Nested door objects inside a room
	Graph-based	RDF, GraphDB	Relationships between building elements

3.2-6 ábra Az adatmodell egy olyan logikai struktúra, amely leírja, hogy az adatok hogyan szerveződnek, tárolódnak és kerülnek feldolgozásra egy rendszerben.

A megfelelő formátum kiválasztása a céloktól függ:

- Ha a szűrés és az elemzés sebessége fontos - relációs táblák (SQL, CSV, RDBMS, oszlopos adatbázisok) is megteszik.
- Ha a struktúra rugalmasságára van szükség - jobb a JSON vagy az XML.
- Ha az adatok összetett kapcsolatokkal rendelkeznek - a gráfadatbázisok átláthatóságot és skálázhatóságot biztosítanak.

A klasszikus relációs adatbázisokban (RDBMS) minden egyes entitást (pl. egy ajtót) egy sor, tulajdonságait pedig a táblázat oszlopai reprezentálják. Például az "Ajtók" kategóriába tartozó elemek táblázata tartalmazhatja az azonosító, a magasság, a szélesség, a tűzállóság és a helyiséget jelző helyiségazonosító mezőket (3.2-7. ábra).

A klasszikus relációs adatbázisokban (RDBMS) a kapcsolatok táblázatok formájában jönnek létre, ahol minden rekord egy objektumot, az oszlopok pedig annak paramétereit jelölik. Táblázatos formátumban a projektben az ajtókra vonatkozó adatok így néznek ki, ahol minden sor egy különálló elemet - egy ajtót, annak egyedi azonosítójával és attribútumaival - képvisel, a szobával való kapcsolat pedig a "szoba azonosítója" paraméteren keresztül valósul meg.



Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Ábra 3.2-7 A projekt "Ajtók" kategórijának három elemére vonatkozó információk táblázatos, strukturált formában.

A lazán strukturált formátumokban, mint például a JSON vagy az XML, az adatokat hierarchikus vagy egymásba ágyazott formában tárolják, ahol az elemek más objektumokat tartalmazhatnak, és szerkezetük változhat. Ez lehetővé teszi az elemek közötti összetett kapcsolatok modellezését. A projektben az ajtókra vonatkozó, strukturált formában rögzített hasonló információk (3.2-7. ábra) lazán strukturált formátumban (JSON) úgy vannak ábrázolva (3.2-8. ábra), hogy azok a szobákon (Rooms - ID) belüli, egymásba ágyazott objektumokká válnak, ami logikailag tükrözi a hierarchiát.

```

1 {
2   "Rooms": [
3     {
4       "ID": 101,
5       "Doors": [
6         {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7         {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8       ]
9     },
10    {
11      "ID": 102,
12      "Doors": [
13        {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14      ]
15    }
16  ]
17}
18

```

Ábra 3.2-8 A projekt "Doors" kategóriájának elemeire vonatkozó információk JSON formátumban.

A gráfmodellben az adatokat csomópontok (csúcsok) és a köztük lévő kapcsolatok (élek) formájában ábrázoljuk. Ez lehetővé teszi az objektumok és attribútumai közötti összetett kapcsolatok megjelenítését. A projektben szereplő ajtó- és szobaadatok esetében a grafikus ábrázolása a következő:

- **A csomópontok (nodes)** a fő entitásokat képviselik: szobák (Room 101, Room 102) és ajtók (ID1001, ID1002, ID1003).
- **A bordák (linkek)** az entitások közötti kapcsolatokat mutatják, pl. egy ajtó tartozása egy bizonyos helyiséghöz.
- **Az attribútumok** csomópontokhoz vannak rendelve és tartalmazzák az egység tulajdonságait (magasság, szélesség, ajtók tűzállósága).

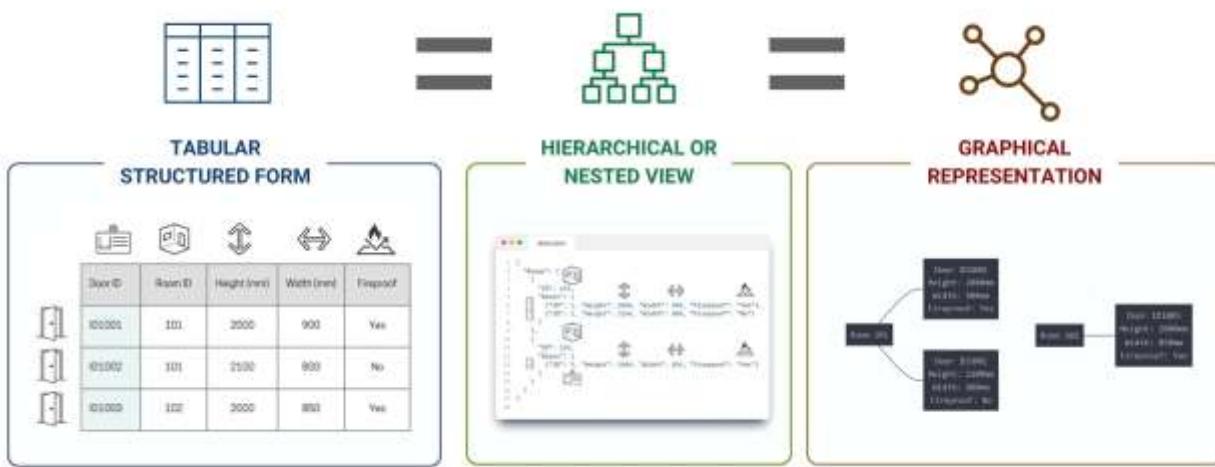


Ábra 3.2-9 A projektkapu entitások információi grafikus nézetben.

Az ajtók leírásának gráf adatmodelljében az egyes szobák és ajtók külön csomópontok. Az ajtók élekkel kapcsolódnak a szobákhoz, amelyek azt jelzik, hogy az ajtó egy adott szobához tartozik. Az ajtók attribútumait (magasság, szélesség, tűzállóság) a megfelelő csomópontok tulajdonságaiként tároljuk. A gráfformátumokról és arról, hogy a gráfszemantika hogyan alakult ki az építőiparban, bővebben a "A szemantika és ontológia kialakulása az építőiparban" című fejezetben lesz szó.

A gráfadatbázisok akkor hatékonyak, ha nem maguk az adatok, hanem az adatok közötti kapcsolatok a fontosak, például az ajánlórendszerekben, útválasztó rendszerekben vagy a létesítménygazdálkodási projektek komplex kapcsolatainak modellezésekor. A gráf formátum leegyszerűsíti az új kapcsolatok létrehozását, mivel lehetővé teszi új adattípusok hozzáadását a gráfhoz a tárolási struktúra megváltoztatása nélkül. A relációs táblázatokhoz és a strukturált formátumokhoz képest azonban a gráfban nincs további adatkapcsolat - a kétdimenziós adatbázis-adatok gráfba való átvitele nem növeli a kapcsolatok számát, és nem nyújt új információt.

Az adatok formáját és sémáját az adott felhasználási esethez és az elvégzendő feladatokhoz kell igazítani. Az üzleti folyamatokban való hatékony munkavégzéshez fontos, hogy olyan eszközöket és olyan adatmodelleket használjunk, amelyek segítségével a lehető leggyorsabban és legegyszerűbben juthatunk eredményre.



ábra 3.2-10 A projektelemekre vonatkozó ugyanazon információk különböző formátumokban, különböző adatmodellek segítségével tárolhatók.

Napjainkban a legtöbb nagyvállalat a túlzottan összetett adatok problémájával szembesül. A több száz vagy ezer alkalmazás mindegyike saját adatmodellt használ, ami túlzott komplexitást eredményez - egy-egy modell gyakran több tucatszor bonyolultabb a szükségesnél, és az összes modell összesítése több ezerszer bonyolultabb. Ez a túlzott összetettség jelentősen megnehezíti mind a fejlesztők, mind a végfelhasználók munkáját.

Ez az összetettség komoly korlátokat szab a vállalat rendszereinek fejlesztésére és karbantartására. A modell minden egyes új eleme további kódot, új logika implementálását, alapos tesztelést és a meglévő megoldásokhoz való hozzáigazítást igényel. Mindez növeli a költségeket és lassítja a vállalat

automatizálási csapatának munkáját, még az egyszerű feladatokat is költséges és időigényes folyamatokká változtatva.

A komplexitás az adatarchitektúra minden szintjét érinti. A relációs adatbázisokban ez az egyre növekvő számú, gyakran redundáns táblában és oszlopban nyilvánul meg. Az objektumorientált rendszerekben a komplexitást az osztályok és az egymással összefüggő tulajdonságok sokasága növeli. Az olyan formátumokban, mint az XML vagy a JSON, a komplexitás zavaros, egymásba ágyazott struktúrákban, egyedi kulcsokban és következetlen sémákban nyilvánul meg.

Az adatmodellek túlzott bonyolultsága nemcsak kevésbé hatékonyá teszi a rendszereket, hanem a végfelhasználók és a jövőben a nagy nyelvi modellek és LLM-ügynökök számára is nehezen érthetővé teszi azokat. Az adatmodellek és az adatfeldolgozás megértésének és összetettségének problémája az, ami felveti a kérdést: hogyan lehet az adatokat elég könnyen használhatóvá tenni ahhoz, hogy azok valóban gyorsan hasznosak legyenek.

Még ha az adatmodelleket bölcsen választják is meg, hasznosságuk drámaian csökken, ha az adatokhoz való hozzáférés korlátozott. A szabadalmaztatott formátumok és zárt platformok akadályozzák az integrációt, bonyolítják az automatizálást, és elveszik a védett információk feletti ellenőrzést, így nem csak egy új adatszilót hoznak létre, hanem egy zárt silót, amelyhez csak a szállító engedélyével lehet hozzáérni. Ahhoz, hogy megértsük a probléma nagyságrendjét, fontos pontosan megvizsgálni, hogy a zárt rendszerek hogyan befolyásolják az építőipar digitális folyamatait.

Szabadalmaztatott formátumok és hatásuk a digitális folyamatokra

A digitalizáció során az építőipari vállalatok egyik legfontosabb kihívása az adatokhoz való korlátozott hozzáférés. Ez megnehezíti a rendszerek integrálását, csökkenti az információk minőségét és megnehezíti a hatékony folyamatok megszervezését. E nehézségek hátterében gyakran a szabadalmaztatott formátumok és zárt szoftvermegoldások használata áll.

Sajnos eddig az építőiparban használt számos program csak azt tette lehetővé, hogy a felhasználó az adatokat védett formátumban vagy felhőalapú tárolóba mentse, amelyhez csak szigorúan korlátozott interfészeken keresztül lehet hozzáérni. És nem ritka, hogy ezek a megoldások a nagyobb gyártók még zártabb rendszereire támaszkodva épülnek. Ennek eredményeként még azok a fejlesztők is, akik nyitottabb architektúrákat szeretnének kínálni, kénytelenek megfelelni a nagy gyártók által diktált szabályoknak.

Míg a modern építőipari adatkezelési rendszerek egyre inkább támogatják a nyílt formátumokat és szabványokat (3.1-5. ábra), a CAD- (BIM)-alapú adatbázisok és a kapcsolódó ERP- és CAFM-rendszerek továbbra is elszigetelt, védett "szigetek" az iparág digitális tájképében (3.2-11. ábra).

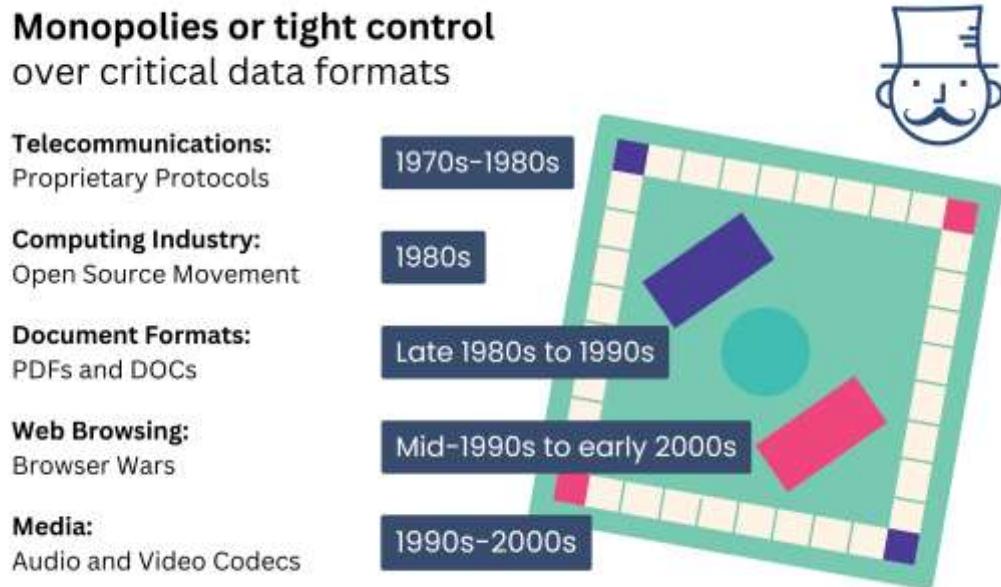


3.2-11 ábra Az adatok zárt és védett jellege akadályokat gördít az adatintegráció és az adathozzáférés előtt.

A zárt és monopolizált formátumok és protokollok nem csak az építőiparban jelentenek problémát. A gazdaság számos ágazatában a zárt szabványok és az adatokhoz való korlátozott hozzáférés elleni küzdelem az innováció lassulásával (3.2-12. ábra), az új szereplők belépésének mesterséges akadályaival és a nagy beszállítótól való függőség elmélyülésével kezdődött. Az adatok jelentőségének gyors növekedésével a versenyhatóságoknak egyszerűen nincs idejük arra, hogy reagáljanak az új digitális piacok által támasztott kihívásokra, és ennek eredményeként a zárt formátumok és az adatokhoz való korlátozott hozzáférés lényegében digitális "határokká" válnak, amelyek korlátozzák az információáramlást és a növekedést [63].

Ha a gépek minden előállítanak, amire szükségünk van, akkor a helyzetünk attól függ, hogy ezeket az árukat hogyan osztják el. Mindenki csak akkor élvezheti a jólétet, ha a gépek által termelt javakat megosztják. Vagy a legtöbb ember mélyszegénységen fog élni, ha az autótulajdonosok sikeresen lobbizhatnak a javak újraelosztása ellen. Eddig úgy tűnik, hogy a dolgok a második irányba haladnak, a technológia egyre nagyobb egyenlőtlenséghöz vezet [72].

- Stephen Hawking, asztrofizikus, 2015



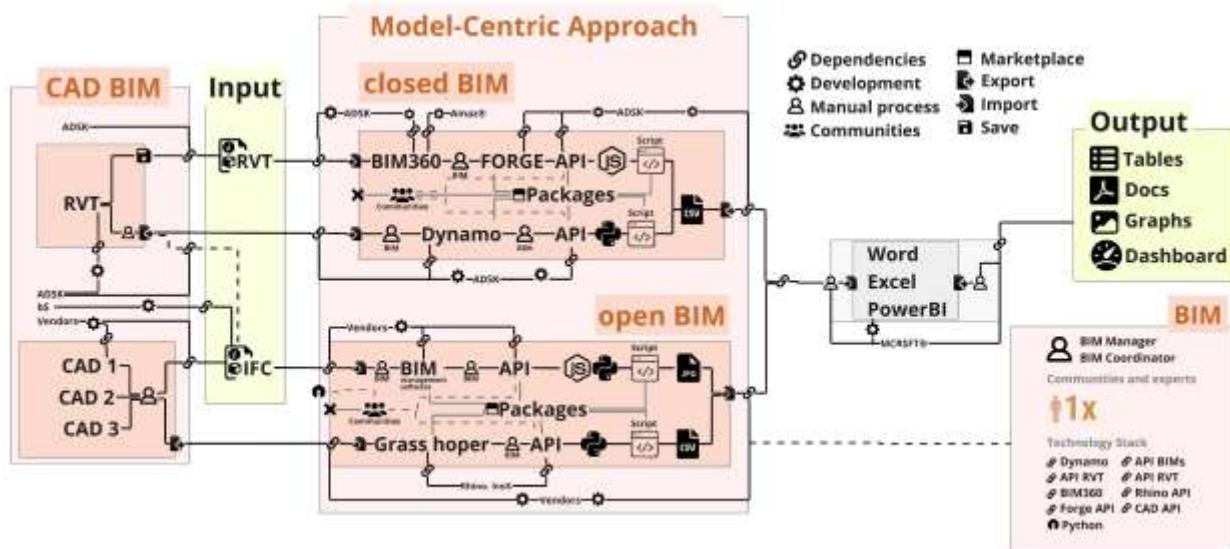
Ábra: 3.2-12 A kulcsfontosságú adatformátumok és protokollok feletti monopólium nem kizárolag az építőipar problémája.

Ennek eredményeképpen az adatbázisokhoz való zárt hozzáférés miatt programok, az adatkezelők, az adatelemzők, az informatikai szakemberek és az építőiparban az adatokhoz való hozzáférésre, feldolgozásra és automatizálásra alkalmazásokat készítő fejlesztők ma számos függőséggel szembesülnek a szoftvergyártótól (3.2-13. ábra). Ezek a függőségek további hozzáférési rétegek formájában speciális API -kapcsolatokkal, valamint speciális eszközökkel és szoftverekkel rendelkező megoldások létrehozását teszik szükségessé.

Az API (Application Programming Interface - alkalmazásprogramozási interfész) egy formalizált interfész, amelyen keresztül az egyik program kölcsönhatásba léphet egy másikkal, adatokat és funkciókat cserélhet anélkül, hogy hozzáférne a forráskódhoz. Az API leírja, hogy egy külső rendszer milyen kéreseket intézhet, milyen formátumban kell azokat benyújtani, és milyen válaszokat kap. Ez egy szabványosított "szerződés" a szoftvermodulok között.

A zárt megoldásoktól való függőségek nagy száma miatt a vállalat teljes kódarchitektúrája és üzleti folyamatlogikája a szoftvergyártónak az adatokhoz való minőségi hozzáférést biztosító politikájától függő eszközök "spaghettiarchitektúrájává" válik.

A zárt megoldásoktól és platformuktól való függés nemcsak a rugalmasság elvesztéséhez, hanem valódi üzleti kockázatokhoz is vezet. A licencfeltételek megváltoztatása, az adatokhoz való hozzáférés lezárása, a formátumok vagy az API-struktúra megváltoztatása - mindenek blokkolhatják a kritikus folyamatokat. Hirtelen kiderül, hogy egy táblázat frissítése integrációk és csatlakozók egész tömbjének átdolgozását igényli (3.2-13. ábra), és a szoftver vagy annak API-szállítójának bármilyen nagyszabású frissítése potenciális veszélyt jelent a vállalat egész rendszerének stabilitására.



ábra 3.2-13 Egy példa a CAD-feldolgozás nagyszámú függőségrére -adatok akadályokat állítanak az adatintegráció elé az építőipari vállalatok ökoszisztémájában.

A fejlesztők és a rendszerarchitektek ilyen körülmények között nem a várakozásért, hanem a túlélésért kénytelenek dolgozni. Új megoldások bevezetése helyett alkalmazkodnak. Fejlesztés helyett a kompatibilitás fenntartására törekednek. A folyamatok automatizálása és felgyorsítása helyett a következő zárt interfések tanulmányozásával, az API dokumentációjával és a végtelen kódújrakészítéssel töltik az idejüket.

A zárt formátumokkal és rendszerekkel való munka nem csupán technikai kihívás, hanem stratégiai kényszer is. A modern automatizálás, az AI, az LLM és a prediktív analitika által kínált nyilvánvaló lehetőségek ellenére sok vállalat nem használja ki a bennük rejlő lehetőségeket. A szabadalmaztatott formátumok (3.2-13. ábra) által felállított akadályok pedig megakadályozzák a vállalkozások hozzáférését saját adataikhoz. Ez talán az építőipar digitális átalakulásának iróniája.

Az adatok átláthatósága és a nyitott rendszerek nem luxus, hanem a gyorsaság és a hatékonyság előfeltételei. Nyitottság nélkül az üzleti folyamatok tele vannak felesleges bürokráciával, többrétegű jóváhagyási láncokkal és a HiPPO elvtől való növekvő függőséggel - a legmagasabb fizetésű személy véleménye alapján hozott döntések.

Mindazonáltal a horizonton paradigmaváltás van kialakulóban. A szabadalmaztatott megoldások dominanciája ellenére egyre több vállalat ismeri fel a negyedik ipari forradalom által inspirált architektúrák korlátait. Ma a vektor az ötödik forradalom elvei felé mozdul el, ahol az adat mint stratégiai eszköz, a nyílt interfések (API-k) és a rendszerek közötti valódi interoperabilitás áll a középpontban.

Ez az átmenet a zárt ökoszisztémáktól a rugalmas, moduláris digitális architektúrák felé való elmozdulást jelenti, ahol a nyílt formátumok, szabványok és az átlátható adatcsere kulcsfontosságúak.

A nyílt formátumok megváltoztatják a digitalizáláshoz való hozzáállást

Az építőipar volt az egyik utolsó, amelyik foglalkozott a zárt és védett adatok problémájával. A gazdaság más ágazataitól eltérően a digitalizáció itt csak lassan fejlődött. Ennek oka többek között az iparág hagyományos konzervatív jellege, a széttagolt helyi megoldások elterjedtsége és a papír alapú dokumentumkezelés mélyen gyökerező jellege. Évtizedeken át a kulcsfontosságú építési folyamatok fizikai rajzokra, telefonthívásokra és szinkronizáltan adatbázisokra támaszkodtak. Ebben az összefüggésben a zárt formátumokat sokáig inkább normának, mint akadálynak tekintették.

Más iparágak tapasztalatai azt mutatják, hogy a zárt adatok előtt álló akadályok megszüntetése az innováció fellendüléséhez, a fejlődés felgyorsulásához és a verseny fokozódásához vezet [73]. A tudományban a nyílt adatok cseréje lehetővé teszi a felfedezések felgyorsítását és a nemzetközi együttműködés fejlesztését. Az orvostudományban javíthatja a diagnózis és a kezelés hatékonyságát. A szoftverfejlesztésben - a közös alkotás és a gyors termékfejlesztés ökoszisztemáinak létrehozását.

A McKinsey "Open Data: Unlock Innovation and Productivity with Information Flow" 2013. [74] a nyílt adatok potenciálisan évente 3-5 millió dollárt szabadíthatnak fel hét kulcsfontosságú iparágban, többek között az építőiparban, a közlekedésben, az egészségügyben és az energetikában. Ugyanezen tanulmány szerint a decentralizált adatökoszisztemák lehetővé teszik a nagy építőipari vállalatok és vállalkozók számára, hogy csökkentsék a szoftverfejlesztési és karbantartási költségeket, felgyorsítva ezzel a digitális bevezetést.

A nyílt architektúrákra való áttérés, amely a gazdaság más ágazataiban már régóta megkezdődött, fokozatosan terjed az építőiparban is. A nagyvállalatok és az állami megrendelők, és különösen az építési projektek beruházásait ellenőrző pénzügyi szervezetek egyre inkább igénylik a nyílt adatok használatát és a számítások, számítások és alkalmazások forráskódjához való hozzáférést. A fejlesztőktől már nem csak azt várják el, hogy digitális megoldásokat hozzanak létre és mutassák meg egy projekt végső számait - elvárják, hogy átláthatóak, reprodukálhatóak és függetlenek legyenek a harmadik fél alkalmazásgyártóktól.

A nyílt forráskódú megoldások használata biztosítja az ügyfél számára, hogy még ha a külső fejlesztők fel is hagynak az együttműködéssel vagy elhagyják a projektet, ez nem befolyásolja az eszközök és rendszerek továbbfejlesztésének lehetőségét. A nyílt adatok egyik fő előnye, hogy az alkalmazásfejlesztők nem függnek az adatok eléréséhez szükséges konkrét platformuktól.

Ha egy vállalat nem tud teljesen lemondani a szabadalmaztatott megoldásokról, egy lehetséges kompromisszum a visszafejtési technikák alkalmazása. Ezek a jogilag és technikailag megalapozott módszerek lehetővé teszik a zárt formátumok átalakítását hozzáférhetőbbé, strukturáltabbá és integrálhatóbbá. Ez különösen akkor fontos, ha szükség van a régebbi rendszerekhez való csatlakozásra vagy az információk egyik szoftverből a másikba történő migrálására.

A nyílt formátumokra való áttérés és a reverse engineering (a védett rendszerek legális feltörése) építőiparban való alkalmazásának egyik legfényesebb példája a számítógépes tervezési rendszerekben (CAD) széles körben használt DWG formátum megnyitásáért folytatott küzdelem története. 1998-ban egy szoftvergyártó monopóliumára válaszul a többi 15 CAD-gyártó új szövetséget hozott létre "Open DWG" néven, hogy a fejlesztőknek ingyenes és független eszközököt biztosítson a DWG formátummal (a rajzátvitel de facto szabványa) való munkához, anélkül, hogy szükségük lenne saját szoftverekre vagy zárt API-kra. Ez az esemény fordulópontot jelentett, amely lehetővé tette, hogy vállalatok tízezrei kapjanak szabad hozzáférést az 1980-as évek végétől napjainkig népszerű CAD-megoldás zárt formátumához, és kompatibilis megoldásokat hozzanak létre, ami elősegítette a versenyt a CAD-piacon [75]. Napjainkban az 1996-ban először létrehozott "Open DWG" SDK-t szinte minden olyan megoldás használja, amelyben a DWG formátum importálása, szerkesztése és exportálása lehetséges, a DWG formátum fejlesztőjének hivatalos alkalmazásán kívül.

Más technológiai óriások is hasonló átalakításokat kényszerítenek ki. A Microsoft, amely egykor a szabadalmaztatott megközelítés szimbóluma volt, megnyitotta a .NET Framework forráskódját, Linuxot kezdett használni az Azure felhőszolgáltatási infrastruktúrájában, és felvásárolta a GitHubot, hogy megerősítse pozícióját a nyílt forráskódú közösségen. [76]. A Meta (korábban Facebook) nyílt forráskódú mesterséges intelligencia modelleket adott ki, például a Llama sorozatot, hogy elősegítse az innovációt és az együttműködést a mesterséges intelligencia-ügynökök fejlesztése terén. Mark Zuckerberg vezérigazgató szerint a következő évtizedben a nyílt forráskódú platformok fognak élen járni a technológiai fejlődésben [77].

A nyílt forráskód egy olyan szoftverfejlesztési és -terjesztési modell, amelyben a forráskód szabadon felhasználható, tanulmányozható, módosítható és terjeszthető.

A nyílt adatok és a nyílt forráskódú megoldások nem csupán trenddé, hanem a digitális fenntarthatóság alapjává válnak. Rugalmasságot, rugalmasságot, a saját döntései feletti ellenőrzést és a digitális folyamatok skálázásának képességét biztosítják a vállalatok számára anélkül, hogy a szállítói irányelvktől függenének. És ami ugyanilyen fontos, visszaadják a vállalkozásoknak az ellenőrzést a 21. század legértékesebb erőforrása - az adataik - felett.

Paradigmaváltás: A nyílt forráskódú mint a szoftvergyártók dominanciájának korszakának vége

Az építőipar olyan változáson megy keresztül, amelyet nem lehet a szokásos módon pénzben kifejezni. Az adatvezérelt, adatközpontú megközelítés koncepciója és a nyílt forráskódú eszközök használata a játékszabályok újragondolásához vezet, amelyen a piac szoftverőriásai állnak.

A korábbi technológiai átalakulásoktól eltérően ezt az átállást a gyártók nem fogják aktívan támogatni. A paradigmaváltás veszélyezteti a hagyományos, licencelésen, előfizetésen és tanácsadáson alapuló üzleti modelljeiket. Az új valóság nem egy kész termékkel vagy fizetős előfizetéssel jár, hanem a folyamatok és a gondolkodásmód átalakítását igényli.

A nyílt technológiákon alapuló adatközponti megoldások kezeléséhez és fejlesztéséhez a vállalatoknak át kell gondolniuk a belső folyamatokat. A különböző részlegek szakembereinek nemcsak együtt kell majd működniük, hanem azt is újra kell gondolniuk, hogyan dolgozzanak együtt.

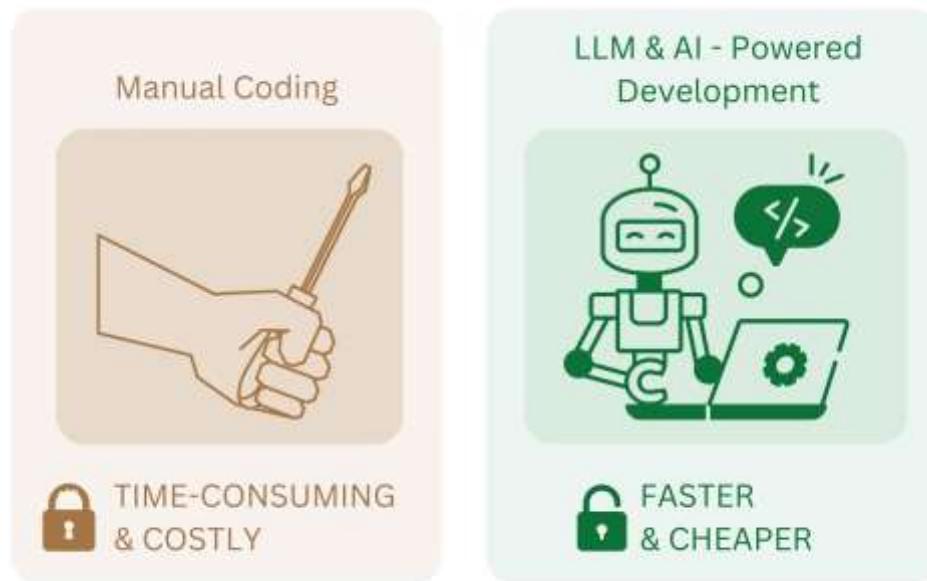
Az új paradigma a nyílt adatok és a nyílt forráskódú megoldások használatát jelenti, ahol a programozók helyett a mesterséges intelligencián és a nagy nyelvi modellekben (LLM) alapuló eszközök játszanak majd különleges szerepet a szoftverkód létrehozásában. Már 2024 közepére a Google-nél az új kód több mint 25 százalékát mesterséges intelligenciával hozzák létre [78]. A jövőben az LLM-ekkel történő kódolás a munka 80 százalékát az idő mindössze 20 százaléka alatt fogja elvégezni (3.2-14. ábra).

A McKinsey 2020 tanulmánya [79] szerint a GPU-k egyre inkább felváltják a CPU-kat az analitikában, mivel nagy teljesítményük és a modern nyílt forráskódú eszközök támogatják őket. Ez lehetővé teszi a vállalatok számára, hogy felgyorsítsák az adatfeldolgozást anélkül, hogy jelentős beruházásokat kellene eszközölniük drága szoftverekbe vagy szűkös szakemberek felvételébe.

Vezető tanácsadó cégek, mint például a McKinsey, a PwC és a Deloitte hangsúlyozzák a nyílt szabványok, a nyílt forráskódú alkalmazások növekvő jelentőségét az iparágakban.

A PwC Open Source Monitor 2019 jelentés [80] szerint a 100 vagy több alkalmazottat foglalkoztató vállalatok 69%-a tudatosan használ nyílt forráskódú megoldásokat. Az OSS-t különösen a nagyvállalatoknál használják aktívan: A 200-499 főt foglalkoztató vállalatok 71%-a, az 500-1999 főt foglalkoztató kategóriában 78%, a 2000 főnél többet foglalkoztató vállalatok körében pedig akár 86%. A Synopsys OSSRA 2023 jelentése szerint az elemzett kódábrisok 96%-a tartalmazott nyílt forráskódú komponenseket [81].

A jövőben a fejlesztő szerepe nem a kézi kódírás lesz, hanem az adatmodellek és áramlási architektúrák megtervezése, valamint a megfelelő számításokat igény szerint létrehozó mesterséges intelligencia-ügynökök irányítása. A felhasználói felületek minimalisztikussá válnak, az interakció pedig párbeszéd-alapúvá. A klasszikus programozás átadja helyét a digitális megoldások magas szintű tervezésének és összehangolásának (3.2-14. ábra). A jelenlegi trendek - mint például a low-code platformok (7.4-6. ábra) és az LLM-alapú ökoszisztemák (7.4-4. ábra) - jelentősen csökkenteni fogják az informatikai rendszerek fejlesztésének és karbantartásának költségeit.



Ábra 3.2-14 Míg ma az alkalmazásokat kézzel készítik a programozók, a jövőben a kód jelentős részét mesterséges intelligencia és LLM-alapú megoldások generálják majd.

Ez az átmenet nem fog hasonlítani a korábbiakhoz, és nem valószínű, hogy a nagy szoftvergyártók lesznek a katalizátorok.

A Harvard Business School "The Value of Open Source Software" 2024 [40] című tanulmányában a nyílt forráskódú szoftverek összértékét két szempontból becsülük meg. Egyrészt, ha kiszámoljuk, hogy mennyibe kerülne az összes létező nyílt forráskódú megoldás nulláról való felépítése, az összeg körülbelül 4,15 milliárd dollár lenne. Másrészt, ha úgy képzeljük el, hogy minden vállalat saját maga fejleszti a nyílt forráskódú megoldások analógjait (ami mindenhol előfordul), anélkül, hogy hozzáférne a meglévő eszközökhöz, akkor a teljes üzleti költség elérné a kolosszális 8,8 millió dollárt - ez a kereslet költsége.

Nem nehéz kitalálni, hogy egyetlen nagy szoftvergyártó sem érdekelte abban, hogy a 8,8 millió dolláros potenciális értékű szoftverpiacot mindössze 4,15 milliárd dollárra zsugorítsa. Ez azt jelentené, hogy a kereslet volumenét több mint 2000-szeresére csökkentenék. Egy ilyen átalakulás egyszerűen veszteséges azon gyártók számára, amelyek üzleti modellje arra épül, hogy évekig fenntartják az ügyfelek zárt megoldásuktól való függőségét. Így azok a vállalatok, amelyek arra számítanak, hogy valaki kényelmes és nyitott, kulcsrakész megoldást kínál nekik, csalódhatnak - ezek a szállítók egyszerűen nem fognak megjelenni.

A nyílt digitális architektúrára való áttérés nem jelent munkahely- vagy bevételkiesést. Éppen ellenkezőleg, megteremti a rugalmas és alkalmazkodó üzleti modellek feltételeit, amelyek végül kiszoríthatják a hagyományos licenc- és dobozos szoftverpiacot.

A licencek eladása helyett szolgáltatások, a zárt formátumok helyett nyílt platformok, a szállítótól való függés helyett függetlenség és a valós igényekre való megoldásépítés képessége. Azok, akik eddig csak használták az eszközöket, képesek lesznek azok társzerzővé válni. Azok pedig, akik adatokkal, modellekkel, forgatókönyvekkel és logikával tudnak dolgozni, az iparág új digitális gazdaságának

középpontjában találják majd magukat. Ezekről a változásokról, valamint arról, hogy milyen új szerepek, üzleti modellek és együttműködési formák alakulnak ki a nyílt adatok körül, bővebben a könyv utolsó, tizedik részében fogunk beszélni.

A nyílt adatokon és nyílt kódon alapuló megoldások lehetővé teszik a vállalatok számára, hogy az üzleti folyamatok hatékonyságára összpontosítsanak, ahelyett, hogy elavult API-kkal és zárt rendszerek integrálásával küzdenének. A nyílt architektúrára való tudatos áttérés jelentősen javíthatja a termelékenységet és csökkentheti a szállítóktól való függőséget.

Az új valóságra való áttérés nemcsak a szoftverfejlesztés megközelítésének megváltozását jelenti, hanem az adatokkal való munka elvének újragondolását is. Ennek az átalakulásnak a középpontjában nem a kód, hanem az információ áll: annak struktúrája, hozzáférhetősége és értelmezhetősége. És itt kerülnek előtérbe a nyílt és strukturált adatok, amelyek az új digitális architektúra szerves részévé válnak.

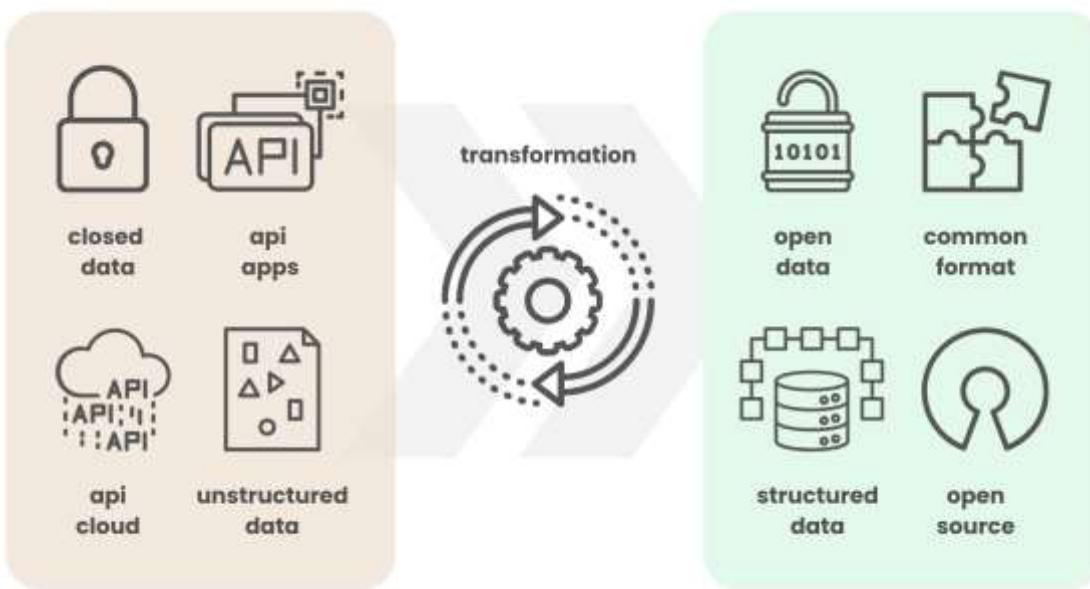
Strukturált nyílt adatok: a digitális átalakulás alapja

Míg az elmúlt évtizedekben az üzleti fenntarthatóságot nagymértékben meghatározta a szoftvermegoldások kiválasztása és a konkrét szállítóktól való függés, addig a mai digitális gazdaságban az adatminőség és az azzal való hatékony munka képessége a kulcstényező. A nyílt forráskód fontos része az új technológiai paradigmának, de a benne rejlő lehetőségek csak akkor szabadulnak fel igazán, ha az adatok érthetők, rendszerezettek és gépileg olvashatóak. Az adatmodellek minden típusa közül a strukturált nyílt adatok válnak a fenntartható digitális átalakulás sarokkövévé.

A strukturált nyílt adatok fő előnye az egyértelmű értelmezés és az automatizált feldolgozás lehetősége. Ez jelentős hatékonyságnövekedést tesz lehetővé mind az egyes műveletek szintjén, mind a szervezet egészének szintjén.

A Deloitte "Az adatátviteli folyamat a vállalati átalakulásban" című jelentése szerint [82] szerint a strukturált adatok átadásának kezelése érdekében az IT-vel való együttműködés kritikus fontosságú. Az Egyesült Királyság kormányának "Data Analytics and AI in Government Project Delivery" (2024) [83] című jelentése szerint az adatmegosztás előtt álló akadályok elhárítása a különböző projektek és szervezetek között kulcsfontosságú a projektmenedzsment hatékonyságának javításához. A dokumentum hangsúlyozza, hogy az adatformátumok szabványosítása és a nyílt adatok elveinek bevezetése elkerülheti az információk megkettőzését, minimalizálhatja az időveszteséget és javíthatja az előrejelzések pontosságát.

Az építőiparban, ahol hagyományosan nagyfokú széttagoltság és a formátumok sokfélesége uralkodik, a strukturált egységesítési folyamat és a strukturált nyílt adatok döntő szerepet játszanak a koherens és kezelhető folyamatok kialakításában (4.1-14. ábra). Lehetővé teszik a projekt résztvevői számára, hogy a zárt platformok, adatmodellek és formátumok közötti inkompatibilitásból adódó technikai problémák megoldása helyett a termelékenység javítására összpontosítsanak.



Ábra 3.2-15 A nyílt strukturált adatok csökkentik a szoftvermegoldásoktól és platformuktól való függőséget, és felgyorsítják az innovációt.

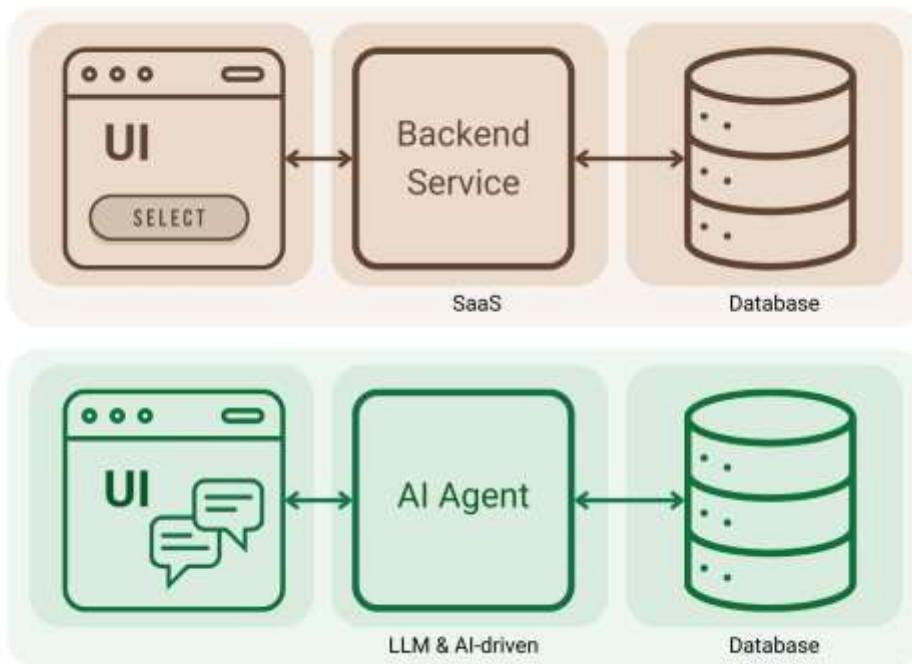
A modern technológiai eszközök, amelyeket a könyv későbbi részében részletesen tárgyalunk, nemcsak az információgyűjtést, hanem annak automatikus tisztítását is lehetővé teszik: a duplikációk kiküszöbölését, a hibák kijavítását és az értékek normalizálását. Ez azt jelenti, hogy az elemzők és mérnökök nem szétszórt dokumentumokkal, hanem egy szervezett, elemzésre, automatizálásra és döntéshozatalra alkalmas tudásbázissal dolgoznak.

Legyen olyan egyszerű, amennyire csak lehet, de ne legyen egyszerűbb.

- Albert Einstein, elméleti fizikus (az idézet szerzőségét vitatják [84]).

Ma már az adatokkal való munkavégzéshez szükséges legtöbb felhasználói felület automatikusan létrehozható - anélkül, hogy minden egyes üzleti esethez kézzel kellene kódot írni. Ehhez olyan infrastrukturális rétegre van szükség, amely további utasítások nélkül megérti az adatszerkezetet, a modellt és a logikát (4.1-15. ábra). A strukturált adatok teszik lehetővé ezt a megközelítést: űrlapok, táblázatok, szűrők és nézetek automatikusan generálhatók minimális programozási erőfeszítéssel.

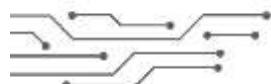
A legfontosabb, a felhasználó számára kritikus interfések még mindig kézi finomítást igényelhetnek. A legtöbb esetben azonban - és ez a munkaforgatókönyvek 50-90 százalékát jelenti - elegendő az alkalmazások és számítások automatikus generálása az erre a céllra szolgáló speciális alkalmazások használata nélkül (3.2-16. ábra), ami jelentősen csökkenti a fejlesztési és karbantartási költségeket, csökkenti a hibákat és felgyorsítja a digitális megoldások megvalósítását.



Ábra 3.2-16 Az adatokkal való munka architektúrális modelljei: hagyományos alkalmazásarchitektúra és AI -orientált modell LLM-mel.

A digitális evolúció következő lépése az egyedi alkalmazásokra épülő architektúrákról a nyelvi modellekkel (LLM) alapuló, intelligensen irányított rendszerek felé történő elmozdulás. Egy ilyen architektúrában a strukturált adatok nem csupán tárolási tényezőként válnak, hanem a kontextus alapján elemezni, értelmezni és cselekvéseket ajánlani képes mesterséges intelligenciaeszközökkel való interakció alapjává is.

A következő fejezetekben a nyílt strukturált adatokon alapuló architektúra megvalósításának valós példáit tekintjük át, és bemutatjuk, hogyan alkalmazzuk a nyelvi modelleket az adatok automatikus értelmezésére, érvényesítésére és feldolgozására. Ezek a gyakorlati esetek segítenek jobban megérteni, hogyan működik az új digitális logika a gyakorlatban - és milyen előnyökkel jár az átalakulásra kész vállalatok számára.



FEJEZET 3.3.

LLM ÉS SZEREPÜK AZ ADATFELDOLGOZÁSBAN ÉS AZ ÜZLETI FOLYAMATOKBAN

LLM csevegőszobák: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok az adatfeldolgozási folyamatok automatizálására.

A nagy nyelvi modellek (Large Language Models, LLM) megjelenése a strukturált nyílt adatok és a nyílt forráskódú filozófia felé irányuló mozgalom természetes folytatása volt. Amikor az adatok rendszerezetté, hozzáférhetővé és gépileg olvashatóvá válnak, a következő lépés egy olyan eszköz, amely képes interakcióba lépni ezekkel az információkkal anélkül, hogy bonyolult kódot kellene írni vagy speciális technikai ismeretekkel kellene rendelkezni.

Az LLM-ek a nyitottság közvetlen termékei: nagyméretű nyílt adathalmazok, publikációk és a nyílt forráskódú mozgalom. Nyílt tudományos cikkek, nyilvánosan elérhető szöveges adatok és a közös fejlesztés kultúrája nélkül nem lenne ChatGPT vagy más LLM. Az LLM bizonyos értelemben az emberiség felhalmozott digitális tudásának "párlata", amelyet a nyitottság elvei alapján gyűjtötték össze és képeztek.

A modern nagy nyelvi modellek (LLM - Large Language Models), mint például a ChatGPT ® (OpenAI), LlaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ a felhasználók számára lehetővé teszik, hogy természetes nyelven fogalmazzanak meg lekérdezéseket az adatokhoz. Ezáltal az információkkal való munka nemcsak a fejlesztők, hanem az elemzők, mérnökök, tervezők, menedzserek és más, a programozástól korábban távol álló szakemberek számára is elérhetővé válik

Az LLM (Large Language Model) egy mesterséges intelligencia, amelyet arra képezték ki, hogy az internetről összegyűjtött hatalmas mennyiségű adat alapján megértse és generálja a szöveget. Képes a kontextus elemzésére, kérdések megválaszolására, párbeszédre, szövegek írására és szoftverkód generálására.

Ha korábban az adatok megjelenítéséhez, feldolgozásához vagy elemzéséhez egy speciális programozási nyelv ismerete volt szükséges: Python, SQL, R vagy Scala, valamint az olyan könyvtárakkal való munka képessége, mint a Pandas, Polars vagy DuckDB és sok más, 2023-tól a helyzet gyökeresen megváltozik. Most már a felhasználó egyszerűen leírhatja, hogy mit szeretne kapni - és a modell maga generálja a kódot, végrehajtja azt, megjelenít egy táblázatot vagy grafikont, és elmagyarázza az eredményt. Évtizedek óta először a technológia fejlődése nem a bonyolítás, hanem a radikális egyszerűsítés és a hozzáférhetőség útját járja.

Ez az elv - "adatok feldolgozása szavakkal (felszólításokkal)". - az információval való munka fejlődésének új szakaszát jelentette, és a megoldások létrehozását az absztrakció még magasabb szintjére emelte. Ahogyan egykor a felhasználóknak már nem volt szükségük az internet technikai alapjainak megértésére ahhoz, hogy online boltokat üzemeltessenek, vagy weboldalakat hozzanak létre a WordPress, a Joomla és más nyílt forráskódú moduláris rendszerek segítségével (av tor a könyv 2005 óta dolgozik ilyen

rendszerrel, többek között oktatási és mérnöki online platformokkal). - Ez viszont a digitális tartalom és az online üzletág fellendüléséhez vezetett - ma már a mérnökök, elemzők és menedzserek programozási nyelvek ismerete nélkül is automatizálhatják a munkafolyamatokat. Ezt nagy teljesítményű LLM-ek segítik elő - ingyenesek és nyílt forráskódúak is, mint például a LLaMA, Mistral, Qwen, DeepSeek és mások -, amelyek a lehető legszélesebb közönség számára teszik elérhetővé a fejlett technológiákat.

Nagy LLM nyelvi modellek: hogyan működik?

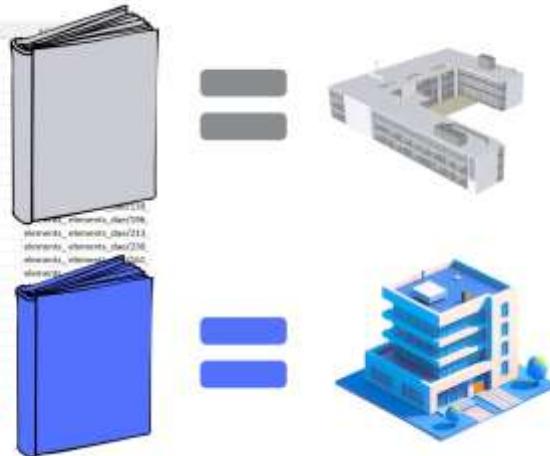
A nagy nyelvi modellek (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) olyan neurális hálózatok, amelyeket az internetről, könyvekből, cikkekből és más forrásokból származó hatalmas mennyiségű szöveges adaton képeznek ki. Fő feladatak az emberi beszéd kontextusának megértése és értelmes válaszok generálása.

A modern LLM a Google kutatói által 2017-ben javasolt Transformer architektúrán alapul [85]. Ennek az architektúrának a kulcskomponense a figyelemmechanizmus, amely lehetővé teszi, hogy a modell figyelembe vegye a szavak közötti kapcsolatokat, függetlenül a szövegenkívül elfoglalt helyüktől.

Az LLM tanulási folyamata távolról is hasonlít ahhoz, ahogyan az emberek nyelvet tanulnak, csak több milliószor nagyobb mértékben. A modell szavak és kifejezések példáinak milliárdjait elemzi, és mintákat azonosít a nyelv szerkezetében és a szemantikai átmenetek logikájában. A teljes szöveget tokenkre - minimális szemantikai egységekre (szavakra vagy azok részeire) - bontja, amelyeket aztán egy többdimenziós térben vektorokká alakít (8.2-2. ábra). Ezek a vektorreprezentációk lehetővé teszik a gép számára, hogy "megértsé" a fogalmak közötti rejtett kapcsolatokat, ahelyett, hogy egyszerűen szimbólumok sorozataként operálna a szöveggel.

A nagy nyelvi modellek nem csak szöveggeneráló eszközök. Képesek felismerni a jelentést, kapcsolatot találni a fogalmak között és dolgozni az adatokkal, még akkor is, ha azok különböző formátumokban vannak megjelenítve. A lényeg az, hogy az információt érhető modellekre kell bontani, és olyan tokenekként kell reprezentálni, amelyekkel az LLM dolgozni tud.

Ugyanez a megközelítés alkalmazható az építési projektekre is. Ha egy projektre egyfajta szövegként gondolunk, ahol minden egyes épület, elem vagy konstrukció egy-egy token, akkor hasonló módon kezdhetjük el feldolgozni az ilyen információkat. Az építési projekteket könyvekhez hasonlíthatjuk, amelyek kategóriákba, fejezetekbe és bekezdéscsoportokba rendeződnek, amelyek minimális tokenekből - az építési projekt elemeiből - állnak (3.3-1. ábra). Az adatmodellek strukturált formátumba való lefordításával a strukturált adatokat vektorbázisokká is lefordíthatjuk (8.2-2. ábra), amelyek ideális forrást jelentenek a gépi tanulás és az olyan technológiák, mint az LLM számára.



Ábra 3.3-1 Az építési projektelem olyan, mint egy szövegben a token: egy minimális egység, amelyből a teljes projekt szakaszainak (kategóriáinak) csoportjai (bekezdések) alakulnak ki.

Ha egy építési projektet digitalizálnak, és annak elemeit tokenek vagy vektorok formájában reprezentálják, akkor a merev formális lekérdezések helyett természetes nyelven lehet hozzáférni hozzájuk. Ez az a pont, ahol az LLM egyik legfontosabb előnye a képbe kerül - a lekérdezés jelentésének megértése és a releváns adatokkal való összekapcsolása.

A mérnöknek többé nem kell SQL -lekerdezést vagy Python kódot írnia a szükséges adatok megszerzéséhez - egyszerűen, az LLM és az adatszerkezet ismeretében a szokásos módon megfogalmazhatja a feladatot: "Keresse meg az összes olyan vasbetonszerkezetet, amelynek betonosztálya magasabb, mint B30, és számíts ki a teljes térfogatukat". A modell felismeri a lekérdezés jelentését, gépileg olvasható formába önti, megkeresi az adatokat (csoportosítja és átalakítja), és visszaadja a végeredményt.

A dokumentumokat, táblázatokat, projektmodelleket vektoros ábrázolásokká alakítják (beágyazás) és az adatbázisban tárolják. Amikor a felhasználó kérdést tesz fel, a lekérdezést szintén vektorba konvertálja, és a rendszer megtalálja a legrelevánsabb adatokat. Ez lehetővé teszi, hogy az LLM ne csak a saját betanított tudására támaszkodjon, hanem a tényleges vállalati adatokra is, még akkor is, ha azok már a modellképzés befejezése után jelentek meg.

Az LLM egyik legfontosabb előnye az építőiparban a programkód generálásának képessége. Ahelyett, hogy a műszaki feladatot átadnák egy programozónak, a szakemberek természetes nyelven leírhatják a feladatot, és a modell létrehozza a szükséges kódot, amelyet (a chatból másolva) felhasználhatnak a folyamatautomatizálási kód létrehozásakor. Az LLM -modellek lehetővé teszik a mély programozási ismeretekkel nem rendelkező szakemberek számára, hogy hozzájáruljanak a vállalat üzleti folyamatainak automatizálásához és fejlesztéséhez.



3.3-2. ábra LLM-ek lehetővé teszik a felhasználók számára, hogy kódot írjanak és eredményeket kapjanak anélkül, hogy programozási ismeretekre lenne szükségük.

A Wakefield Research által végzett és az SAP által szponzorált 2024-es tanulmány [36] szerint, amely 300, legalább 1 milliárd dolláros éves bevétellel rendelkező amerikai vállalat felsővezetőjét kérdezte meg: a felsővezetők 52%-a bízik a mesterséges intelligenciában az adatok elemzésében és a döntéshozatalhoz szükséges ajánlásokban. További 48% használja a mesterséges intelligenciát korábban nem számolt kockázatok azonosítására, 47% pedig arra, hogy alternatív terveket javasoljon. Emellett 40 százalékuk új termékfejlesztésre, költségvetés-tervezésre és piackutatásra használja a mesterséges intelligenciát. A tanulmány kimutatta az AI pozitív hatását a magánéletre is: a válaszadók 39%-a számolt be a munka és a magánélet jobb egyensúlyáról, 38%-a a mentális egészség javulásáról, 31%-a pedig a stressz alacsonyabb szintjéről.

Az LLM-ek azonban minden erejük ellenére olyan eszköz maradnak, amelyet fontos tudatosan használni. Mint minden technológiának, ezeknek is vannak korlátaik. Az egyik legismertebb probléma az úgynevezett "hallucinációk" - olyan esetek, amikor a modell magabiztosan egy hihető, de ténytől eltérő téves választ ad. Ezért kritikus fontos megérteni, hogyan működik a modell: milyen adatokat és adatmodellekkel képes hiba nélkül értelmezni, hogyan értelmezi a lekérdezéseket, és honnan szerzi az információkat. Azt is érdemes szem előtt tartani, hogy az LLM ismeretei a képzésének időpontjára korlátozódnak, és külső adatokkal való kapcsolat nélkül a modell nem feltétlenül veszi figyelembe az aktuális normákat, szabványokat, árakat vagy technológiákat.

A megoldás ezekre a problémákra a vektoradatbázisok rendszeres frissítése, a releváns forrásokhoz való kapcsolódás és az autonóm AI -ügynökök kifejlesztése, amelyek nem csak kérdésekre válaszolnak, hanem proaktívan használják az adatokat a képzéshez, kezelik a feladatokat, azonosítják a kockázatokat, optimalizálási lehetőségeket javasolnak és figyelemmel kísérik a projekt teljesítményét.

Az LLM -felületekre való áttérés az építőiparban nem csupán technológiai újdonság. Ez egy paradigmaváltás, amely megszünteti az emberek és az adatok közötti akadályokat. Arról van szó, hogy az információkkal olyan könnyen tudunk dolgozni, mint ahogyan egymással beszélünk - és mégis pontos, ellenőrzött és megvalósítható eredményeket kapunk.

Azok a vállalatok, amelyek korábban kezdték el használni ezeket az eszközöket, mint mások, jelentős

versenyelőnyre tesznek szert. Ez magában foglalja a munka felgyorsítását, a költségek csökkentését és a tervezési megoldások minőségének javítását az adatalemzéshez való gyors hozzáférés és az összetett kérdésekre adott gyors válaszok megtalálásának képessége révén. De biztonsági kérdéseket is figyelembe kell venni. A felhőalapú LLM -szolgáltatások használata az adatszivárgás kockázatával járhat. Ezért a szervezetek egyre inkább olyan alternatív megoldásokat keresnek, amelyek lehetővé teszik számukra, hogy az LLM-eszközöket saját infrastruktúrájukban - helyben, az információk teljes körű védelme és ellenőrzése mellett - telepítsék.

Helyi LLM-ek használata az érzékeny vállalati adatok esetében

Az első chat-LLM-ek megjelenése 2022-ben új szakaszt jelentett a mesterséges intelligencia fejlődésében. Közvetlenül e modellek széles körű elterjedése után azonban felmerült egy jogos kérdés: mennyire biztonságos a vállalati adatok és lekérdezések felhőbe történő átvitele? A legtöbb felhőalapú nyelvi modell a kommunikációs előzményeket és a feltöltött dokumentumokat a szervereiken tárolta, és az érzékeny információkkal foglalkozó vállalatok számára ez komoly akadályt jelentett a mesterséges intelligencia bevezetésének.

Az egyik legtartósabb és leglogikusabb megoldás erre a problémára a nyílt forráskódú LLM helyben, a vállalati informatikai infrastruktúrán belül történő telepítése. A felhőszolgáltatásokkal ellentétben a helyi modellek internetkapcsolat nélkül működnek, nem továbbítják az adatokat külső szerverekre, és a vállalatok teljes ellenőrzést biztosítanak az információik felett

A legjobb nyílt modell [Open Source LLM] jelenleg hasonló teljesítményt nyújt, mint a zárt modellek [például ChatGPT, Claude], de körülbelül egy év lemaradással [77].

- Ben Cottier, az Epoch AI, egy nonprofit kutatószervezet vezető kutatója, 2024

A nagy technológiai vállalatok elkezdték az LLM-jeiket helyi használatra elérhetővé tenni. A Meta nyílt forráskódú LLaMA-sorozata és a Kínából származó, gyorsan növekvő DeepSeek projekt volt a példa a nyílt architektúra felé való elmozdulásra. Mellettük a Mistral és a Falcon is nagy teljesítményű, a szabadalmaztatott platformok korlátaitól mentes modelleket adott ki. Ezek a kezdeményezések nem csupán felgyorsították a globális mesterséges intelligencia fejlesztését, hanem valódi alternatívákat is nyújtottak az adatvédelemre odafigyelő vállalatoknak a függetlenség, a rugalmasság és a biztonsági megfelelés terén.

A vállalati környezetben, különösen az építőiparban, az adatvédelem nem csupán kényelmi szempont, hanem a jogszabályoknak való megfelelés kérdése is. A pályázati dokumentumokkal, becslésekkel, rajzokkal és bizalmas levelezéssel való munka szigorú ellenőrzést igényel. És ez az a terület, ahol a helyi LLM biztosítja a szükséges biztosítékokat arra, hogy az adatok a vállalat határain belül maradjanak.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
Data Control	Data is transmitted to third parties	Data remains within the company's network
License	Proprietary, paid	Open-source (Apache 2.0, MIT)
Infrastructure	Requires internet	Operates in an isolated environment
Customization	Limited	Full adaptation to company needs
Cost	Pay-per-token/request	One-time hardware investment + maintenance costs
Scalability	Easily scalable with cloud resources	Scaling requires additional local hardware
Security & Compliance	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
Performance & Latency	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
Integration	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
Updates & Maintenance	Automatically updated by provider	Requires manual updates and model retraining
Energy Consumption	Energy cost is covered by provider	High power consumption for inference and training
Offline Availability	Not available without an internet connection	Works completely offline
Inference Cost	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Ábra 3.3-3 A helyi modellek teljes körű ellenőrzést és biztonságot, míg a felhőalapú megoldások egyszerű integrációt és automatikus frissítéseket kínálnak.

A helyi nyílt forráskódú LLM:

- Teljes ellenőrzés az adatok felett. minden információ a vállalon belül marad, kizárva az illetéktelen hozzáférést és az adatszivárgást.
- Offline működés. Nem függ az internetkapcsolattól, ami különösen fontos az elszigetelt IT-infrastruktúrák esetében. Ez szankciók vagy blokkolt felhőszolgáltatások esetén is biztosítja a zavartalan működést.
- Alkalmazási rugalmasság. A modell használható szöveggenerálásra, adatelemzésre, programkód írására, tervezési támogatásra és üzleti folyamatok kezelésére.
- Alkalmazkodás a vállalati célkitűzésekhez. Az LLM belső dokumentumokon képezhető, ami lehetővé teszi a vállalati munka sajátosságainak és iparági jellemzőinek figyelembevételét. A helyi

LLM összekapcsolható a CRM, az ERP vagy a BI platformokkal, lehetővé téve az ügyfélkérdések elemzésének automatizálását, a jelentések készítését vagy akár a trendelőrejelzést.

A DeepSeek ingyenes és nyílt forráskódú modelljének -R1-7B telepítése egy szerveren, egy egész felhasználói csapat számára, havi 1000 dolláros költséggel, potenciálisan kevesebb kerülhet, mint a felhő API-k, mint például a ChatGPT vagy a Claude éves díjai, és lehetővé teszi a vállalatok számára, hogy teljes mértékben átvegyék az adataik feletti ellenőrzést, megszünteti azok internetre történő továbbítását, és segít megfelelni az olyan szabályozási követelményeknek, mint a GDPR

Más iparágakban a helyi LLM-ek már megváltoztatják az automatizáláshoz való hozzáállásukat. A támogató szolgáltatásokban a gyakori ügyfélkérdésekre válaszolnak, csökkentve az operátorok munkaterhelését. A HR-osztályokon elemzik az önéletrajzokat és kiválasztják a megfelelő jelölteket. Az e-kereskedelemben személyre szabott ajánlatokat generálnak a felhasználói adatok felfedése nélkül.

Hasonló hatás várható az építőipari ágazatban is. Az LLM projektadatokkal és szabványokkal való integrációjának köszönhetően fel lehet gyorsítani a dokumentáció elkészítését, automatizálni lehet a becslések és a prediktív költségelemzések elkészítését. Az LLM strukturált táblázatokkal és adatkeretekkel együtt történő használata különösen igéretes területté válik.

Az AI teljes ellenőrzése a vállalaton belül és a saját LLM bevetésének módja

A modern eszközök lehetővé teszik a vállalatok számára, hogy néhány óra alatt helyben telepítsenek egy nagy nyelvi modellt (LLM). Ez teljes ellenőrzést biztosít az adatok és az infrastruktúra felett, kiküszöbölte a külső felhőszolgáltatásoktól való függőséget és minimalizálva az információszivárgás kockázatát. Ez a megoldás különösen fontos az érzékeny projektdokumentációval vagy kereskedelmi szempontból érzékeny adatokkal dolgozó szervezetek számára.

A feladataktól és erőforrásuktól függően különböző telepítési forgatókönyvek állnak rendelkezésre, a kész megoldásuktól a rugalmásabb és skálázhatóbb architektúráig. Az egyik legegyszerűbb eszköz az Ollama, amely lehetővé teszi a nyelvi modellek szó szerint egyetlen kattintással történő futtatását, mély technikai tudás nélkül. Gyors kezdés az Ollamával:

1. Töltsé le a disztribúciót az operációs rendszeréhez (Windows / Linux / macOS) a hivatalos weboldalról: ollama.com
2. Telepítse a modellt a parancssoron keresztül. Például a *Mistral* modell esetében:

```
ollama run mistral
```

3. A modell futtatása után a modell készen áll a munkára - a terminálon keresztül küldhet szöveges lekérdezéseket, vagy integrálhatja más eszközökbe. Futtassa a modellt és hajtson végre egy lekérdezést:

ollama run mistral "Hogyan lehet létrehozni egy számítást az összes erőforrással a 100 mm széles gipszkarton válaszfal telepítéséhez szükséges munkához?".

Azok számára, akik inkább egy ismerős vizuális környezetben szeretnének dolgozni, ott van az LM Studio, egy ingyenes alkalmazás, amelynek felülete a ChatGPT-re emlékeztet

- Telepítse az LM Studio-t a disztribúciós készlet letöltésével a hivatalos weboldalról - lmstudio.ai
- A beépített katalóguson keresztül válasszon ki egy modellt (pl. Falcon vagy GPT-Neo-X) és töltse le.
- A modellel való munka a ChatGPT intuitív, de teljesen lokalizált felületén keresztül történik.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

3.3-4. ábra A népszerű helyi nyílt forráskódú LLM -modellek összehasonlítása.

A modell kiválasztása a sebességgel, pontossággal és a rendelkezésre álló hardveres lehetőségekkel szemben támasztott követelményektől függ (3.3-4. ábra). Az olyan kis modellek, mint a Mistral 7B és a Baichuan 7B alkalmasak könnyű feladatokra és mobil eszközökre, míg az olyan nagy teljesítményű modellek, mint a DeepSeek -V3 jelentős számítási erőforrásokat igényelnek, de nagy teljesítményt és több nyelv támogatását kínálják. Az elkövetkező években az LLM-piac gyorsan fog növekedni - egyre több könnyű és speciális modellt fogunk látni. Az általános célú LLM-ek helyett, amelyek az összes emberi tartalmat lefedik, szűk szakterületi szakértelemre kiképzett modellek fognak megjelenni. Például olyan modellek

megjelenésére számíthatunk, amelyeket kizárálag mérnöki számítások, építési becslések vagy CAD-formátumú adatok kezelésére terveztek. Az ilyen specializált modellek gyorsabbak, pontosabbak és biztonságosabbak lesznek - különösen olyan szakmai környezetekben, ahol fontos a nagy megbízhatóság és a tárgyi mélység.

A helyi LLM elindítása után a vállalat sajátos feladataihoz igazítható. Erre a célra a finomhangolási technikát alkalmazzák, amelynek során a modellt továbbképzik belső dokumentumokon, műszaki utasításokon, szerződéssablonokon vagy projektdokumentáción.

RAG: Intelligens LLM - a vállalati adatokhoz hozzáférő asszisztensek

Az LLM üzleti alkalmazás fejlődésének következő lépcsőfoka a modellek integrálása a tényleges, valós idejű vállalati adatokkal. Ezt a megközelítést RAG-nak (Retrieval-Augmented Generation) - Visszakeresés és kiegészített generálás - nevezik. Ebben az architektúrában a nyelvi modell nem csupán egy párbeszédes felületté válik, hanem egy teljes értékű intelligens asszisztenssé, amely képes navigálni a dokumentumokban, rajzokban, adatbázisokban és pontos, kontextusfüggő válaszokat adni.

A RAG fő előnye, hogy a modell előzetes betanítása nélkül képes a vállalat belső adatait felhasználni, miközben az információkezelés nagy pontossága és rugalmassága megmarad.

A RAG technológia két fő összetevőt egyesít:

- **Lekérdezés:** a modell csatlakozik az adattárolókhhoz - dokumentumok, táblázatok, PDF -fájlok, rajzok - és a felhasználó kérésének megfelelően lekérdezi a releváns információkat.
- **Kiegészített generálás:** a modell a kinyert adatok alapján pontos, megalapozott választ generál, figyelembe véve a lekérdezés kontextusát és specifikusságát.

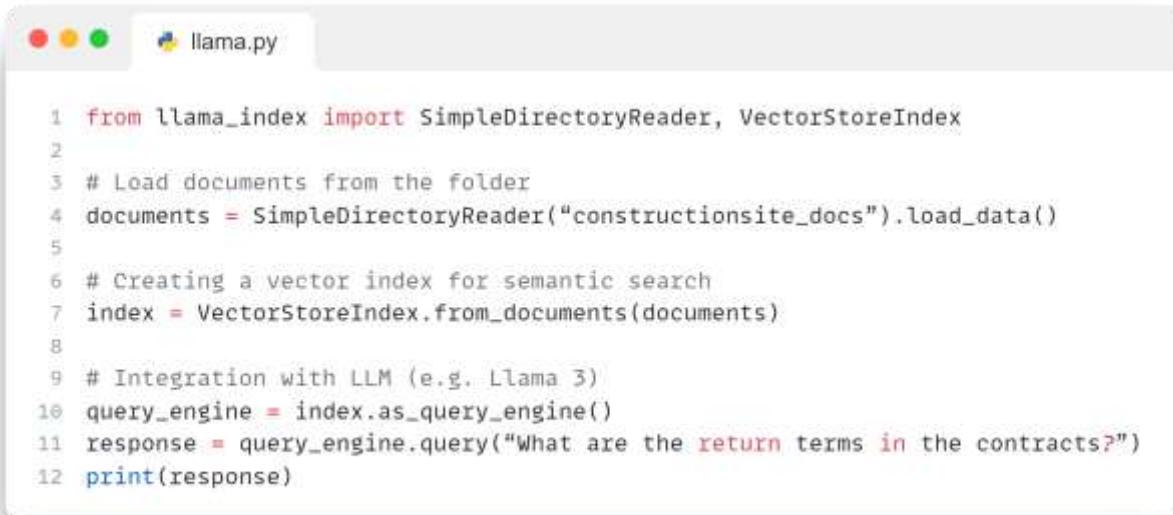
Az LLM RAG támogatással történő futtatásához, néhány lépést kell követnie:

- **Adatelőkészítés:** a szükséges dokumentumok, rajzok, specifikációk, táblázatok összegyűjtése. Ezek különböző formátumúak és szerkezetűek lehetnek, a PDF-től az Excelig.
- **Indexelés és vektorizálás:** olyan eszközökkel, mint a Llmalndex vagy a LangChain, az adatokat vektoros reprezentációkká alakítjuk át, amelyek lehetővé teszik a szövegrészletek közötti szemantikai kapcsolatok megtalálását (a vektoros adatbázisokról és a nagy tömbök vektoros reprezentációjáról, beleértve a CAD projekteket is, bővebben a 8. részben).
- **Kérdezze meg az asszisztentet:** az adatok feltöltése után kérdéseket tehet fel a modellnek, amely a címen keresi meg a válaszokat a vállalati keretek között, nem pedig az internetről gyűjtött általános ismeretekben.

Tegyük fel, hogy egy vállalatnak van egy constructionsite_docs mappája, ahol a szerződéseket, utasításokat, becsléseket és táblázatokat tárolja. Egy Python szkript segítségével (3.3-5. ábra) beolvashatjuk ezt a mappát, és vektoros indexelést építhetünk: minden dokumentumot a szöveg szemantikai tartalmát tükröző vektorok halmazává alakítunk át. Ezáltal a dokumentumok egyfajta "jelentés-térképpé" válnak, amelyen a modell hatékonyan tud navigálni, és kapcsolatot találni a kifejezések

és mondatok között.

A modell például "emlékszik" arra, hogy a "visszaküldés" és a "reklamáció" szavak gyakran szerepelnek a szerződésnek az építési helyszínre történő anyagszállításra vonatkozó részében. Ezután, ha feltesznek egy kérdést - például: "Mennyi a visszaküldési időnk?". (3.3-5. ábra - a kód 11. sora) - az LLM elemzi a belső dokumentumokat, és megtalálja a pontos információt, intelligens asszisztensként viselkedve, amely képes elolvasni és megérteni az összes vállalati fájl tartalmát.



```

1 from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3 # Load documents from the folder
4 documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6 # Creating a vector index for semantic search
7 index = VectorStoreIndex.from_documents(documents)
8
9 # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)

```

Ábra 3.3-5 Az LM beolvas egy fájlokat tartalmazó mappát - hasonlóan ahhoz, ahogyan az ember megnyitja azt, és megkeresi a kívánt dokumentumot

A kód bármely számítógépen futtatható, ahol a Python telepítve van. A Python és az IDE-k használatáról a kód futtatásához a következő fejezetben fogunk bővebben beszélni.

Az LLM helyi telepítése nem csupán egy trend, hanem stratégiai megoldás a biztonságot és rugalmasságot értékelő vállalatok számára. Az LLM, akár helyi vállalati számítógépeken, akár online megoldások segítségével történő telepítése azonban csak az első lépés. Ahhoz, hogy az LLM képességeit valós feladatokra alkalmazhassák, a vállalatoknak olyan eszközöket kell használniuk, amelyek lehetővé teszik számukra, hogy ne csak a chat-válaszokat fogadják, hanem a létrehozott logikát kód formájában tárolják, amely az LLM használatának kontextusán kívül is futtatható. Ez fontos a megoldások skálázása szempontjából - a megfelelően szervezett folyamatok lehetővé teszik, hogy a mesterséges intelligencia fejlesztéseket egyszerre több projektben vagy akár az egész vállalaton belül alkalmazzák.

Ebben az összefüggésben fontos szerepet játszik a megfelelő fejlesztőkörnyezet (IDE) kiválasztása. A modern programozóeszközök nem csak LLM-alapú megoldások fejlesztését teszik lehetővé, hanem azok integrálását a meglévő üzleti folyamatokba is, automatizált ETL -Pipeline -vé alakítva azokat



FEJEZET 3.4.

IDE AZ LLM TÁMOGATÁSSAL ÉS JÖVŐBELI PROGRAMOZÁSI VÁLTOZÁSOKKAL

Az IDE kiválasztása: az LLM kísérletekből az üzleti megoldásokig

Az automatizálás, az adatalemzés és a mesterséges intelligencia világában való elmerüléskor - különösen, ha nagy nyelvi modellekkel (LLM) dolgozunk - kritikus fontosságú a megfelelő integrált fejlesztőkörnyezet (IDE) kiválasztása. Ez az IDE lesz az Ön fő munkaeszköze: az a hely, ahol az LLM által generált kód futtatható lesz, mind a helyi számítógépen, mind a vállalati hálózaton belül. Az IDE kiválasztása nemcsak a munka kényelmét határozza meg, hanem azt is, hogy milyen gyorsan tud majd a kísérleti LLM-lekérdezésektől a valós üzleti folyamatokba ágyazott, teljes értékű megoldásokig eljutni.

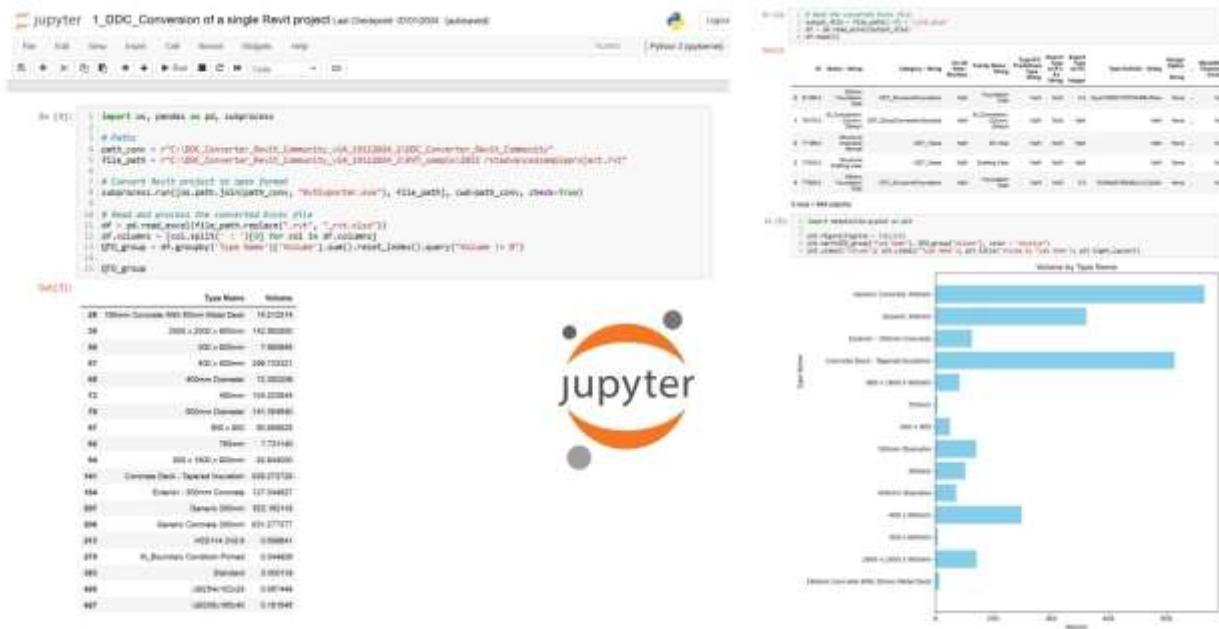
Az IDE (integrált fejlesztőkörnyezet) egy sokoldalú építőelem a számítógépen a folyamatautomatizáláshoz és az adatfeldolgozáshoz. Ahelyett, hogy külön-külön tartana fűrészt, kalapácsot, fúrót és egyéb szerszámokat, egyetlen eszközzel rendelkezik, amely minden tud: vágni, rögzíteni, fúrni és még az anyagok minőségét is ellenőrizheti. Az IDE a programozók számára egy olyan egységes tér, ahol kódot írhat (az építkezés analógiájára - rajzokat készíthet), tesztelheti a munkáját (épületmodell összeszerelése), megtalálhatja a hibákat (mint az építkezés során a szerkezetek szilárdságának ellenőrzése) és futtathatja a kész projektet (a ház üzembe helyezése).

A népszerű IDE-k áttekintése:

- A **PyCharm®** (JetBrains) egy nagy teljesítményű professzionális IDE a Pythonhoz. A nagyszámú beépített funkcióval köszönhetően jól használható komoly projektekhez. Az interaktív Jupyter-fájlok (IPYNB) alapvető támogatása azonban csak a fizetős változatban érhető el, és a kezdők számára a felület túlterhelő lehet.

Az IPYNB kiterjesztésű fájl (Interactive Python Notebook) az interaktív Jupyter® Notebooks (3.4-1. ábra) formátuma, amelyben a kód, a vizualizáció és a magyarázat egyetlen dokumentumban egyesül. Ez a formátum ideális jelentések, analitikai és képzési forgatókönyvek készítéséhez.

- A **VS Code®** (Microsoft) egy gyors, rugalmas és testreszabható eszköz ingyenes IPYNB-támogatással és számos bővítménnyel. Kezdők és profik számára egyaránt alkalmas. Lehetővé teszi a GitHub Copilot és a nyelvi modell bővítmények integrálását, így kiváló választás AI és adattudományi projektekhöz.
- **Jupyter Notebook** - Klasszikus és népszerű választás kísérletezéshez és tanuláshoz. Lehetővé teszi a kód írását, magyarázatok hozzáadását és az eredmények vizualizálását egyetlen felületen (3.4-1. ábra). Ideális hipotézisek gyors teszteléséhez, LLM munkához és reprodukálható vad adatalemzési lépések létrehozásához. A függőségek és könyvtárak kezeléséhez az Anaconda Navigator használatát javasoljuk, amely egy vizuális felület a Python környezet kezelésére.



Ábra 3.4-1 Jupyter Notebook a Pipeline folyamatok létrehozásának egyik legkényelmesebb és legnépszerűbb eszköze.

■ **A Google Collab™** (és a Kaggle platform (9.2-5. ábra)) a Jupyter felhőalapú alternatívája, amely ingyenes GPU/TPU hozzáférést biztosít. Ez egy nagyszerű megoldás a kezdéshez - nincs helyi szoftver telepítése és közvetlenül egy böngészőből dolgozhatunk vele. Támogatja az integrációt a Google Drive-val és újabban a Geminivel (a Google LLM).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

3.4-2 IDE összehasonlítás: Jupyter Notebook az egyik legkényelmesebb és legegyszerűbb eszköz a Pipeline folyamatok létrehozásához.

Az IDE kiválasztása a feladatoktól függ. Ha gyorsan szeretne elkezdeni dolgozni az AI, próbálja ki a Jupyter Notebookot vagy a Google Collabot. Komoly projektekhez jobb a PyCharm vagy a VS Code használata. A lényeg az, hogy elkezdjük. A modern eszközök lehetővé teszik, hogy a kísérletekből gyorsan működő megoldások szülessenek.

A fent ismertetett IDE-k mindegyike lehetővé teszi adatfeldolgozási csővezetékek létrehozását, azaz kódblokk-modulok láncolatát (amelyeket az LLM generálhat), amelyek mindegyike például egy-egy szakaszért felelős:

- elemzési forgatókönyvek,
- a dokumentumokból történő információszerzés láncai,
- automatikus válaszok a RAG alapján,
- jelentések és vizualizációk készítése.

A moduláris felépítésnek köszönhetően minden lépés külön blokkként ábrázolható: adatbetöltés→ szűrés→ elemzés→ vizualizáció→ eredmények exportálása. Ezek a blokkok újrafelhasználhatók, - átalakíthatók és új láncokba illeszthetők, mint egy konstruktur, csak az adatokhoz.

A mérnökök, menedzserek és elemzők számára ez megnyitja a döntéshozatali logika dokumentálásának lehetőségét az LLM-mel generálható kód formájában. Ez a megközelítés segít felgyorsítani a rutinfeladatokat, automatizálni a tipikus műveleteket és olyan megismételhető folyamatokat létrehozni, amelyekben minden lépés egyértelműen dokumentált és átlátható a csapat minden tagja számára.

Az automatizált ETL Pipelines (7.2-3. ábra), Apache Airflow (7.4-4. ábra), Apache NiFi (7.4-5. ábra) és n8n (7.4-6. ábra) eszközök a folyamatok automatizálásához szükséges logikai blokkok felépítésére a könyv 7. és 8. részében részletesebben tárgyaljuk.

IDE LLM-támogatással és jövőbeli programozási változtatások

A mesterséges intelligencia integrálása a fejlesztési folyamatokba megváltoztatja a programozási környezetet. A modern környezetek már nem csupán szövegszerkesztők szintaxis-kiemeléssel - intelligens asszisztensekké válnak, amelyek képesek megérteni a projekt logikáját, kiegészíteni a kódot, és akár meg is magyarázzák, hogyan működik egy adott kód részlet. Olyan termékek jelennek meg a piacon, amelyek a mesterséges intelligenciát a hagyományos fejlesztés határainak feszetegetésére használják:

- **GitHub Copilot** (integrálható VS Code, PyCharm): mesterséges intelligencia -asszisztens, amely a megjegyzések vagy részleges leírások alapján kódot generál, és a szöveges tippeket kész megoldásokká alakítja.
- **Cursor** (a VS Code elágazása AI -kernel): nem csak a kódírás befejezését teszi lehetővé, hanem azt is, hogy kérdéseket tegyen fel a projektnek, függőségeket keressen és tanuljon a kódbazisból.
- **JetBrains AI Assistant**: a JetBrains IDE (beleértve a PyCharmot is) bővítménye, amelynek funkciója az összetett kód magyarázata, optimalizálás és tesztkészítés.
- **Amazon CodeWhisperer**: a Copilot analógja, amely a biztonságra és az Amazon AWS-szolgáltatások támogatására összpontosít.

A programozás drámai változásra megy keresztül az elkövetkező években. A fő hangsúly a rutinszerű kódírásról a modell- és adatarchitektúra-tervezésre fog áthelyeződni - a fejlesztők jobban részt vesznek majd a rendszertervezésben, míg az AI átveszi a sablonfeladatokat: kódgenerálás, tesztek, dokumentáció és alapvető funkciók. A programozás jövője az emberek és a mesterséges intelligencia együttműködése, ahol a gépek átveszik a technikai rutint, az emberek pedig a kreativitásra koncentrálnak.

A természetes nyelvű programozás minden napos lesz. Az IDE személyre szabása új szintre lép - a

fejlesztőkörnyezetek megtanulnak alkalmazkodni a felhasználó munkastílusához és vállalataihoz azáltal, hogy előre látják a mintákat, kontextuális megoldásokat kínálnak és tanulnak a korábbi projektekből

Ez nem szünteti meg a fejlesztő szerepét, de radikálisan átalakítja azt: a kódírás helyett a tudás, a minőség és a folyamatok irányítása lesz a feladata. Ez az evolúció az üzleti intelligenciát is érinti, ahol a jelentések, vizualizációk és döntéstámogató alkalmazások létrehozása egyre inkább a kód és a logika generálásával történik majd a mesterséges intelligencia és az LLM, a chat és az ügynöki interfések segítségével.

Miután egy vállalat létrehozta az LLM-csevegéseket és kiválasztotta a megfelelő fejlesztési környezetet, a következő fontos lépés az adatok rendszerezése. Ez a folyamat magában foglalja az információk kinyerését a különböző forrásokból, tisztítását, strukturált formába való átalakítását és a vállalati rendszerekbe való integrálását.

Az adatkezelés modern adatközpontú megközelítésében az egyik fő cél az adatok egységes, univerzális formába hozása, amely számos eszközzel és alkalmazással kompatibilis. A strukturálási folyamatok és a strukturált adatok kezeléséhez speciális könyvtárakra van szükség. Az egyik legerősebb, lerugalmásabb és legnépszerűbb a Pandas könyvtár a Python számára. Lehetővé teszi a táblázatos adatok kényelmes feldolgozását: szűrést, csoportosítást, tisztítást, csatolást, aggregálást és jelentést.

Python Pandas: nélkülözhetetlen eszköz az adatokkal való munkához.

A Pandas különleges helyet foglal el az adatalemzés és automatizálás világában. A Python programozási nyelv egyik legnépszerűbb és legszélesebb körben használt könyvtára [86], amelyet strukturált adatokkal való munkára terveztek.

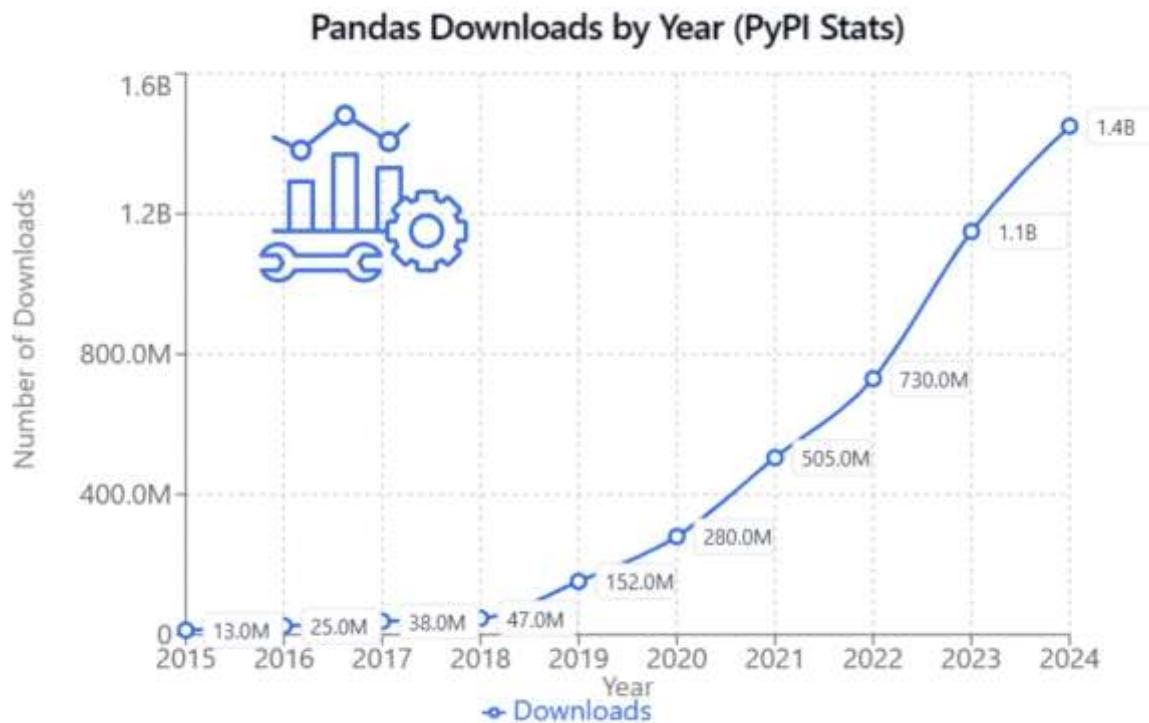
A könyvtár olyan, mint egy kész eszközkészlet: függvények, modulok, osztályok. Ahogyan egy építkezésen sem kell minden alkalommal kalapácsot vagy szintezőt feltalálni, úgy a programozásban a könyvtárak lehetővé teszik, hogy gyorsan megoldjuk a problémákat anélkül, hogy újra feltalálnánk az alapvető funkciókat és megoldásokat.

A Pandas egy nyílt forráskódú Python könyvtár, amely nagy teljesítményű és intuitív adatszerkezeteket biztosít, különösen a DataFrame, egy univerzális formátumot a táblázatokkal való munkához. A Pandas svájci bicska az adatokkal dolgozó elemzők, mérnökök és fejlesztők számára.

A Python egy egyszerű szintaxis-sel rendelkező, magas szintű programozási nyelv, amelyet aktívan használnak az analitikában, az automatizálásban, a gépi tanulásban és a webfejlesztésben. Népszerűségét a kód olvashatóságának, platformokon átívelő jellegének és a könyvtárak gazdag ökoszisztemájának köszönheti. A mai napig több mint 137 000 nyílt forráskódú csomagot hoztak létre Pythonhoz [87], és ez a szám szinte naponta növekszik. minden ilyen könyvtár egyfajta kész funkciók tárháza: az egyszerű matematikai műveletektől a képfeldolgozás, a nagy adatalemzés, a neurális hálózatok és a külső szolgáltatásokkal való integráció komplex eszközeiig.

Más szóval, képzelje el, hogy ingyenes és nyílt hozzáférése van több százezer kész szoftvermegoldáshoz - könyvtárakhoz és eszközökhöz, amelyeket közvetlenül beágyazhat üzleti folyamataiba. Olyan ez, mint egy hatalmas katalógus az automatizáláshoz, elemzéshez, vizualizáláshoz, integráláshoz és még sok másra szükséges alkalmazásokból - és mindenekkel a Python telepítése után azonnal elérhetőek.

A Pandas az egyik legnépszerűbb csomag a Python ökoszisztémában. A Pandas könyvtár átlagos letöltésszáma 2022-ben elérte a napi 4 millió letöltést (3.4-3. ábra), míg 2025 elejére ez a szám napi 12 millió letöltésre nőtt, ami tükrözi a növekvő népszerűséget és széles körű használatát az adatelemzésben és az LLM chat [86]



Ábra 3.4-3 Pandas az egyik legtöbbet letöltött könyvtár. Éves letöltéseinek száma 2024-ben meghaladta az 1,4 milliárdot.

A Pandas könyvtár lekérdezési nyelve funkcionálisában hasonlít az SQL lekérdezési nyelvhez, amelyet a "Relációs adatbázisok és az SQL lekérdezési nyelv" című fejezetben tárgyalunk.

Az analitika és a strukturált adatkezelés világában a Pandas kiemelkedik egyszerűségével, gyorsaságával és teljesítményével, és számos eszközt kínál a felhasználóknak az információk hatékony elemzéséhez és feldolgozásához.

Mindkét eszköz - az SQL és a Pandas - erőteljes adatmanipulációs képességeket biztosít, különösen a

hagyományos Excelhez képest. Támogatják az olyan műveleteket, mint a kiválasztás, szűrés (3.4-4. ábra), a különbség csupán annyi, hogy az SQL a relációs adatbázisokkal való munkára optimalizált, míg a Pandas az adatokat RAM-ban dolgozza fel, ami lehetővé teszi, hogy bármely számítógépen fussen, anélkül, hogy adatbázisokat kellene létrehozni és külön infrastruktúrát telepíteni.

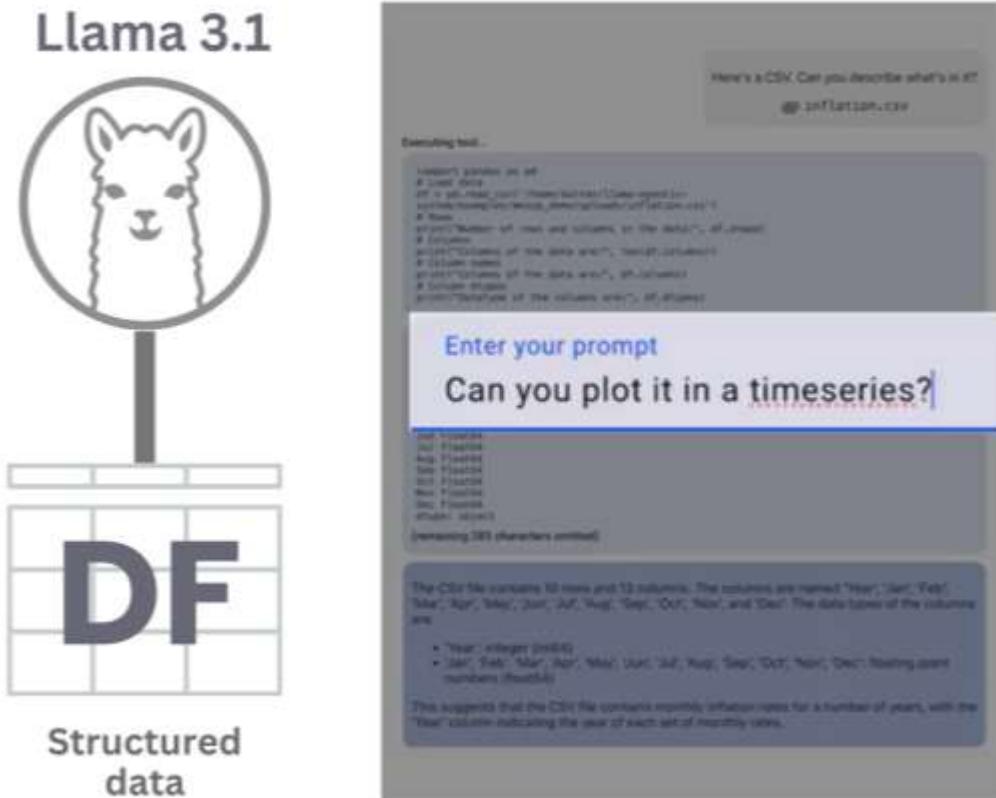


Ábra 3.4-4 A Pandas az SQL -tól eltérően rugalmasan képes különböző adatformátumokkal dolgozni, nem korlátozódva az adatbázisokra.

A Pandas gyakran előnyben részesül tudományos kutatáshoz, folyamatautomatizáláshoz, pipelinek létrehozásához (beleértve az ETL) és adatmanipulációhoz a Python, míg az SQL egy adatbázis-kezelési szabvány, és gyakran használják vállalati környezetben nagy mennyiségű adat kezelésére.

A Python programozási nyelv Pandas könyvtára lehetővé teszi, hogy ne csak olyan alapvető műveleteket hajtson végre, mint a táblázatok olvasása és írása, hanem összetettebb feladatokat is, például adatok összevonását, adatok csoportosítását és összetett analitikus számítások elvégzését.

A Pandas könyvtárat ma már nemcsak a tudományos kutatásban és az üzleti analitikában használják, hanem az LLM -modellekkel együtt is. A Meta® divízió (Facebook™) például az új nyílt forráskódú LLaMa 3.1 modell 2024-es kiadásakor különös figyelmet fordított a strukturált adatokkal való munkára, és a kiadás egyik legfontosabb és első esetének pontosan a CSV formátumú strukturált adatkeretek (3.4-5. ábra) feldolgozását tette meg és a Pandas könyvtárral való integrációt közvetlenül a chatben.



Ábra 3.4-5 A Meta csapat egyik első és legfontosabb esete, amelyet az LlaMa 3.1-ben 2024-ben mutatott be, a Pandas segítségével történő alkalmazásépítés volt.

A Pandas alapvető eszköz több millió adattudós számára, akik adatokat dolgoznak fel és készítenek elő a generatív mesterséges intelligencia számára. A Pandas felgyorsítása nulla kódváltatással hatalmas előrelépés lesz. Az adattudósok órák helyett percek alatt feldolgozhatják az adatokat, és nagyságrendekkel több adatot kapnak a generatív AI modellek képzéséhez [88].

- Jensen Huang, az NVIDIA alapítója és vezérigazgatója

A Pandas segítségével olyan adathalmazok kezelése és elemzése lehetséges, amelyek messze meghaladják az Excel képességeit. Míg az Excel jellemzően legfeljebb 1 millió sornyi adat kezelésére képes, addig a Pandas könnyedén kezeli a több tízmillió sort tartalmazó adathalmazokat (9.1-2. ábra, 9.1-10. ábra) [89]. Ez a képesség lehetővé teszi a felhasználók számára, hogy kifinomult adatelemzést és vizualizációt végezzenek nagy adathalmazokon, mély betekintést nyújtva és megkönyítve az adatvezérelt döntéshozatalt. A Pandas emellett erős közösségi támogatással rendelkezik [90]: világszerte több százmillió fejlesztő és elemző (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) használja online vagy offline nap mint nap, így számos kész megoldást kínál bármilyen üzleti problémára.

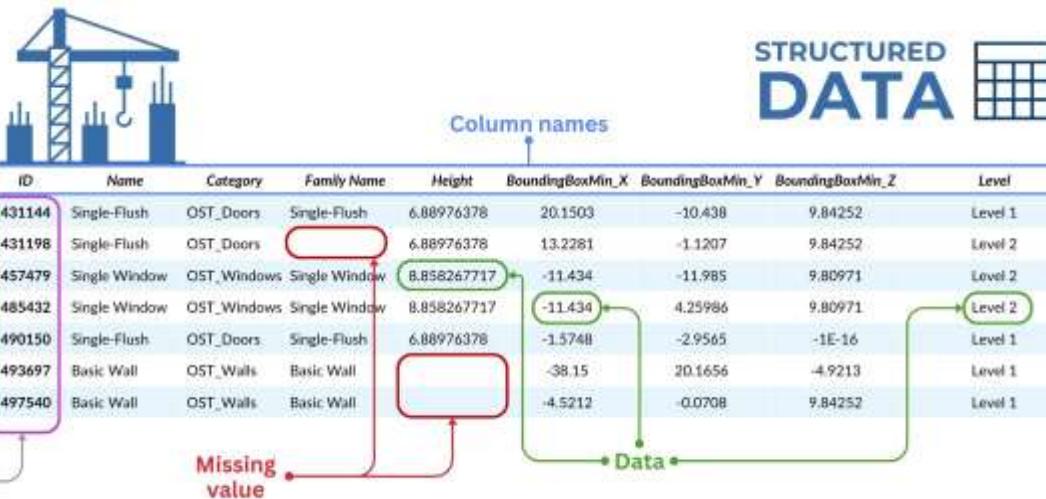
A legtöbb Python analitikai folyamat középpontjában egy strukturált adatforma áll, az úgynevezett DataFrame, amelyet a Pandas könyvtár biztosít. Ez egy hatékony és rugalmas eszköz a táblázatos adatok rendszerezésére, elemzésére és vizualizálására.

DataFrame: univerzális táblázatos adatformátum

A DataFrame a Pandas könyvtár központi struktúrája, amely egy kétdimenziós táblázat (3.4-6. ábra), ahol a sorok az egyes objektumoknak vagy rekordoknak, az oszlopok pedig azok jellemzőinek, paramétereinek vagy kategóriáinak felelnek meg. Ez a struktúra vizuálisan hasonlít az Excel táblázatokra, de rugalmasságban, méretezhetőségeben és funkcionálisban sokkal jobb.

A DataFrame a számítógép RAM-jában tárolt táblázatos adatok ábrázolásának és feldolgozásának egy módja.

A DataFrame a számítógép RAM-jában tárolt táblázatos adatok ábrázolásának és feldolgozásának egy módja. Egy táblázatban a sorok tükrözhetik például egy építési projekt elemeit, az oszlopok pedig azok tulajdonságait: kategóriákat, méreteket, koordinákat, költségeket, feltételeket és így tovább. Ráadásul egy ilyen táblázat egyaránt tartalmazhat egyetlen projektre vonatkozó információkat (4.1-13. ábra) és több ezer különböző projektből származó több millió objektumra vonatkozó adatokat (9.1-10. ábra). A vektorizált Pandas műveleteknek köszönhetően az ilyen mennyiségű információ nagy sebességgel történő szűrése, csoportosítása és összesítése egyszerű.



The diagram illustrates the structure of a DataFrame using a construction site metaphor. A blue crane is positioned above a table. The table has 'Column names' at the top and 'Index label' on the left. The columns are labeled 'ID', 'Name', 'Category', 'Family Name', 'Height', 'BoundingBoxMin_X', 'BoundingBoxMin_Y', 'BoundingBoxMin_Z', and 'Level'. The rows are indexed from 431144 to 497540. Arrows point from the labels to their respective parts in the table. A red box highlights a missing value '8.88247717' in the 'Family Name' column. A green box highlights a row with multiple missing values. A green bracket on the right side of the table is labeled 'Data'. A legend on the left defines the colors: blue for columns, purple for index labels, red for missing values, and green for data.

ID	Name	Category	Family Name	Height	BoundingBoxMin_X	BoundingBoxMin_Y	BoundingBoxMin_Z	Level
431144	Single-Flush	OST_Doors	Single-Flush	6.88976378	20.1503	-10.438	9.84252	Level 1
431198	Single-Flush	OST_Doors		6.88976378	13.2281	-1.1207	9.84252	Level 2
457479	Single Window	OST_Windows	Single Window	8.88247717	-11.434	-11.985	9.80971	Level 2
485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971	Level 2
490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.5748	-2.9565	-1E-16	Level 1
493697	Basic Wall	OST_Walls	Basic Wall		-38.15	20.1656	-4.9213	Level 1
497540	Basic Wall	OST_Walls	Basic Wall		4.5212	0.0708	9.84252	Level 1

3.4-6. ábra Építési projekt mint adatkeret A egy kétdimenziós táblázat, amelynek elemei sorokban, attribútumai pedig oszlopokban találhatók.

Az Nvidia becslése szerint már ma is az összes számítási erőforrás akár 30 százalékát strukturált adatok - adatképek - feldolgozására használják, és ez az arány folyamatosan növekszik.

Az adatfeldolgozás az, amit valószínűleg a világ számítástechnikájának egyharmada végez minden vállalatnál. A legtöbb vállalat adatfeldolgozása és adatai DataFrame, táblázatos formátumban vannak

- Jensen Huang, az Nvidia vezérigazgatója [91]

Soroljuk fel a DataFrame néhány fő jellemzőjét a Pandas:

- **Oszlopok:** a DataFrame oldalon az adatok oszlopokba vannak rendezve, amelyek mindegyike egyedi névvel rendelkezik. Az attribútumoszlopok különböző típusú adatokat tartalmazhatnak, hasonlóan az adatbázisok oszlopaihoz vagy a táblázatok oszlopaihoz.
- **Pandas A sorozat** egy egydimenziós adatszerkezet a Pandasban, hasonlóan egy lista vagy oszlop egy táblázatban, ahol minden érték egy másik indexnek felel meg.

A Pandas sorozat több mint 400 attribútummal és módszerrel rendelkezik, ami hihetetlenül rugalmassá teszi az adatokkal való munkát. Közvetlenül alkalmazhatja a négy száz rendelkezésre álló függvény egyikét egy oszlopra, matematikai műveleteket végezhet, adatokat szűrhet, értékeket helyettesíthet, dátumokkal, karakterláncokkal dolgozhat és még sok minden mást. A Series emellett támogatja a vektorizált műveleteket, ami a ciklikus számításokhoz képest jelentősen felgyorsítja a nagy adathalmazok feldolgozását. Például könnyedén megszorozhatja az összes értéket egy számmal, pótolhatja a hiányzó adatokat, vagy bonyolult átalakításokat alkalmazhat bonyolult ciklusok írása nélkül.

- **Sorok:** a DataFrame-ben a egyedi értékekkel indexelhető. Ez az index lehetővé teszi az adatok gyors módosítását és kiigazítását az egyes sorokban.
- **Index:** Alapértelmezés szerint egy DataFrame létrehozásakor Pandas minden sorhoz hozzárendel egy indexet 0 és N-1 között (ahol N az összes sor száma a DataFrame-ben). Az index azonban módosítható, hogy speciális jelöléseket, például dátumokat vagy egyedi jellemzőket tartalmazzon.
- A sorok **indexelése** egy DataFrame-ben azt jelenti, hogy minden sorhoz egyedi nevet vagy címkét rendelünk, amelyet DataFrame indexnek nevezünk.
- **Adattípusok:** A DataFrame számos adattípust támogat, többek között: A szöveges adatokhoz a `int`, `float`, `bool`, `datetime64` és `object`. minden DataFrame oszlopnak saját adattípusa van, amely meghatározza, hogy milyen műveletek végezhetők a tartalmával.
- **Adatműveletek:** A DataFrame támogatja az adatfeldolgozási műveletek széles skáláját, beleértve az aggregálást ('groupby'), az egyesítést ('merge' és 'join'), az összekapcsolást ('concat'), az osztás-applikáció-kombinációt és sok más adattranszformációs technikát.
- **Méretmanipuláció:** A DataFrame lehetővé teszi oszlopok hozzáadását és eltávolítását, valamint a sorok hozzáadását és eltávolítását, így a struktúra dinamikusan, az adatelemzési igényeknek megfelelően módosítható.
- **A adatok vizualizálása:** a beépített vizualizációs technikák vagy a népszerű adatvizualizációs könyvtárakkal, például a Matplotlib vagy a Seaborn segítségével a DataFrame könnyen átalakítható grafikonokká és diagramokká az adatok grafikus bemutatása érdekében.
- **Adatbevitel és -kimenet:** A Pandas funkciókat biztosít az adatok olvasásához, importálásához és

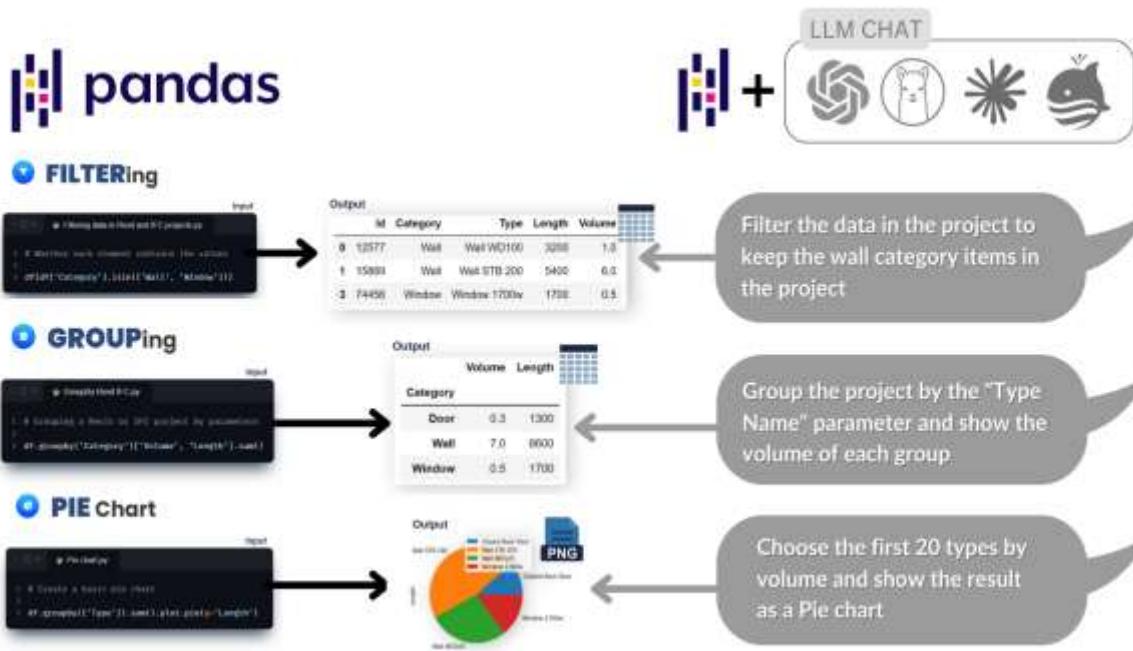
exportálásához különböző fájiformátumokba, például CSV, Excel, JSON, HTML és SQL, ami a DataFrame adatgyűjtés és -elosztás központi központjává teheti.

A CSV és az XLSX -vel ellentétben a Pandas DataFrame nagyobb rugalmasságot és teljesítményt biztosít az adatokkal való munka során: nagy mennyiségű információt tud kezelni a RAM-ban, támogatja a kiterjesztett adattípusokat (beleértve a dátumokat, logikai értékeket és időszorokat), és kiterjedt képességeket biztosít az adatok szűrésére, összesítésére, összevonására és megjelenítésére. Míg a CSV nem tárol információt az adattípusokról és az adatszerkezetről, az XLSX pedig gyakran túlterhelt a formázással, és alacsony a skálázhatósága, a DataFrame továbbra is optimális választás a gyors analitikához, a folyamatok automatizálásához és a mesterséges intelligencia -modellekkel való integrációhoz (3.4-7. ábra). A következő fejezetekben az adatok minden egyes ilyen aspektusát részletesen megvizsgáljuk, továbbá a könyv 8. részében részletesen tárgyaljuk a hasonló formátumokat, mint a Parquet, Apache Orc, JSON, Feather, HDF5 és az adattárházak (8.1-2. ábra).

		XLSX	CSV	Pandas DataFrame
	Storage	Tabular	Tabular	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
	Compression	Built-in	None	None (in-memory)
	Performance	Low	Medium	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Low
	Data Type Support	Limited	Very limited	Extended
	Scalability	Low	Low	Medium (memory limited)

Ábra 3.4-7 DataFrame az optimális választás a nagy teljesítményű és fejlett adattípus-támogatással rendelkező adatmanipulációhoz.

A Pandas könyvtár és a DataFrame formátum rugalmasságának, teljesítményének és egyszerű használatának köszönhetően a Python adatelemzés de facto szabványává vált. Ideálisak mind egyszerű jelentések készítéséhez, mind összetett analitikai pipelinek építéséhez, különösen LLM modellekkel együtt.



Ábra 3.4-8 LLM-ek leegyszerűsítik a Pandas-szal való interakciót: kód helyett elegendő egy szöveges lekérdezés.

Ma a Pandast aktívan használják az LLM-alapú csevegőszobákban, mint például a ChatGPT, LLaMa, DeepSeek, QWEN és mások. Sok esetben, amikor egy modell táblázatfeldolgozással, adatérvényesítéssel vagy analitikával kapcsolatos lekérdezést kap, pontosan a Pandas könyvtár segítségével generál kódot. Ezáltal a DataFrame természetes "nyelvű" válik az adatok mesterséges intelligenciával folytatott párbeszédeken történő ábrázolásához (3.4-8. ábra).

Az olyan modern adattechnológiák, mint a Pandas, megkönnyítik az adatok elemzését, automatizálását és integrálását az üzleti folyamatokba. Gyorsan szolgáltatnak eredményeket, csökkentik a szakemberek munkaterhelését és biztosítják a megismételhető műveleteket.

Következő lépések: fenntartható adatkeret kialakítása

Ebben a részben megnéztük az építőiparban használt legfontosabb adattípusokat, megismertük a tárolásukhoz használt különböző formátumokat, és elemeztük a modern eszközök, köztük az LLM és az IDE-k szerepét az információk feldolgozásában. Megtanultuk, hogy a hatékony adatkezelés a megalapozott döntéshozatal és az üzleti folyamatok automatizálásának alapja. Azok a szervezetek, amelyek képesek strukturálni és rendszerezni adataikat, jelentős versenyelőnyre tesznek szert az adatfeldolgozás és - átalakítás szakaszaiban.

Összefoglalva ezt a részt, érdemes kiemelni a főbb gyakorlati lépéseket, amelyek segítenek a tárgyalt megközelítések alkalmazásában a napi feladatok során:

- Végezze el folyamatainak adatauditját
- Készítsen leltárt a projektjeiben használt összes adattípusról.

- Határozza meg, hogy mely adattípusok és modellek a legkritikusabbak az üzleti folyamatok szempontjából.
- Azonosítani azokat a problémás területeket, ahol az információk gyakran strukturálatlanok, rosszul strukturáltak vagy hozzáférhetetlenek.
- Kezdje el az adatkezelési stratégia kialakítását**
 - A különböző adattípusok kezelésére vonatkozó szakpolitikai kérdések és szabványok felvetése
 - Elemezze, hogy mely munkafolyamatok javíthatók a strukturálatlan adatok strukturált adatokká alakításával.
 - Hozzon létre olyan adattárolási és hozzáférési szabályzatot, amely figyelembe veszi a biztonságot és a titkosságot.
- Az adatokkal való munkavégzéshez szükséges alapvető eszközök telepítése és elsajátítása**
 - Válassza ki a feladatainak megfelelő IDE-t (pl. telepítse a VS Code-ot vagy a Jupyter Notebookot).
 - Próbáljon meg létrehozni egy helyi LLM címet, hogy bizalmasan kezelje személyes adatait.
 - Kísérletezés a Pandas könyvtárral XLSX táblázatos adatok feldolgozásához
 - Írja le az LLM-nek, hogy milyen tipikus feladatokat végez táblázatkezelő eszközökben vagy adatbázisokban, és kérje meg az LLM-et, hogy automatizálja a munkát a Pandas segítségével.

Ezeknek a lépéseknek az alkalmazása lehetővé teszi, hogy fokozatosan átalakítsa az adatokkal való munkavégzéshez való hozzáállását, és a széttagolt, strukturálatlan adathalmazokból egységes ökoszisztemává váljon, ahol az adatok hozzáférhető és érthető eszközévé válnak. Kezdje kicsiben - hozza létre az első DataFrame-et a Pandasban, futtasson egy helyi LLM-et, automatizálja az első rutinfeladatát Python használatával (pl. Excel táblázatok).

A könyv negyedik része az adatminőségre, az adatok szervezésére, strukturálására és modellezésére összpontosít. Olyan módszerekre fogunk összpontosítani, amelyek a különböző adatforrásokat - a PDF-ektől és szövegektől a képeig és CAD-modellekig - strukturált, elemzésre és automatizálásra alkalmas adathalmazokká alakítják át. Azt is megvizsgáljuk, hogyan formalizálják az adatkövetelményeket, hogyan épülnek fel fogalmi és logikai modellek az építési projektekben, és hogyan segíthetnek ebben a folyamatban a modern nyelvi modellek (LLM).





IV RÉSZ

ADATMINŐSÉG: SZERVEZÉS, STRUKTURÁLÁS, MODELLEZÉS

A negyedik rész a széttagolt információk strukturált és jó minőségű adathalmazokká történő átalakításának módszereire és technológiáira összpontosít. Részletesen tárgyalja az építési projektek hatékony információs architektúrájának alapjául szolgáló adatkövetelmények kialakításának és dokumentálásának folyamatait. Bemutatásra kerülnek a strukturált információk különböző forrásokból (PDF -dokumentumok, képek, szöveges fájlok, CAD - modellek) történő kinyerésének gyakorlati módszerei és a megvalósítás példái. Elemezik a reguláris kifejezések (RegEx) és más eszközök használatát az adatok automatikus érvényesítésére és ellenőrzésére. Az adatmodellezés folyamatát koncepcionális, logikai és fizikai szinten lépésről lépésre ismertetjük, figyelembe véve az építőipar sajátosságait. Konkrét példákat mutatunk be a nyelvi modellek (LLM) használatára az információk strukturálási és hitelesítési folyamatainak automatizálására. Az elemzési eredmények vizualizálására hatékony megközelítéseket javasolnak, növelve az analitikus információk hozzáférhetőségét az építési projektmenedzsment minden szintjén

FEJEZET 4.1.

AZ ADATOK STRUKTURÁLT FORMÁBA TÖRTÉNŐ ÁTALAKÍTÁSA

Az adatvezérelt gazdaság korában az adatok a döntéshozatal alapjává válnak, nem pedig akadályává. Ahelyett, hogy az információkat folyamatosan az egyes új rendszerekhez és azok formátumaihoz igazítanák, a vállalatok egyre inkább arra törekszenek, hogy egyetlen strukturált adatmodellt alakítsanak ki, amely az igazság univerzális forrásaként szolgál minden folyamat számára. A modern információs rendszereket nem a formátumok és interfészük, hanem az adatok jelentése köré tervezik - mert a struktúra változhat, de az információ jelentése sokkal tovább marad ugyanaz.

Az adatokkal való hatékony munka kulcsa nem az adatok végtelen átalakításában és átalakításában rejlik, hanem a kezdetektől fogva a megfelelő rendszerezésben: egy olyan univerzális struktúra létrehozásában, amely a projekt életciklusának minden szakaszában képes átláthatóságot, automatizálást és integrációt biztosítani.

A hagyományos megközelítés minden egyes új platform bevezetésekor manuális kiigazításokat tesz szükséges: az adatok migrálása, az attribútumnevek megváltoztatása és a formátumok kiigazítása. Ezek a lépések nem javítják az adatok minőségét, hanem csak elfedik a problémákat, és a végtelen átalakítások örödigi körét hozzák létre. Ennek eredményeként a vállalatok függővé válnak az adott szoftvermegoldásuktól, és a digitális átalakulás lelassul.

A következő fejezetekben megvizsgáljuk, hogyan strukturáljuk helyesen az adatokat, majd hogyan hozzunk létre univerzális modellek, hogyan minimalizáljuk a platformfüggőséget, és hogyan összpontosítsunk arra, ami a legfontosabb - az adatokra mint stratégiai erőforrásra, amelyre fenntartható folyamatok épülnek.

Dokumentumok, PDF, képek és szövegek strukturált formátummá alakításának megtanulása.

Az építési projektekben az információk túlnyomó többsége strukturálatlan formában létezik: műszaki dokumentumok, munkárészek, rajzok, műszaki leírások, ütemtervezek, jegyzőkönyvek. Ezek sokfélése - mind formátumuk, mind tartalmuk tekintetében - megnehezíti az integrációt és az automatizálást.

A strukturált vagy félig strukturált formátumokba történő átalakítás folyamata a bemeneti adatok típusától és a kívánt feldolgozási eredményektől függően változhat.

Az adatok strukturálatlan formából strukturált formába történő átalakítása egyszerre művészettel és tudománnyal. Ez a folyamat a bemeneti adatok típusától és az elemzés céljától függően változik, és gyakran az adatmérnök (3.2-5. ábra) és az elemző munkájának jelentős részét veszi igénybe, és célja egy tiszta, rendezett adathalmaz előállítása.



Ábra: 4.1-1 Struktúrálatlan szkennelt dokumentum átalakítása strukturált táblázatos formátumba.

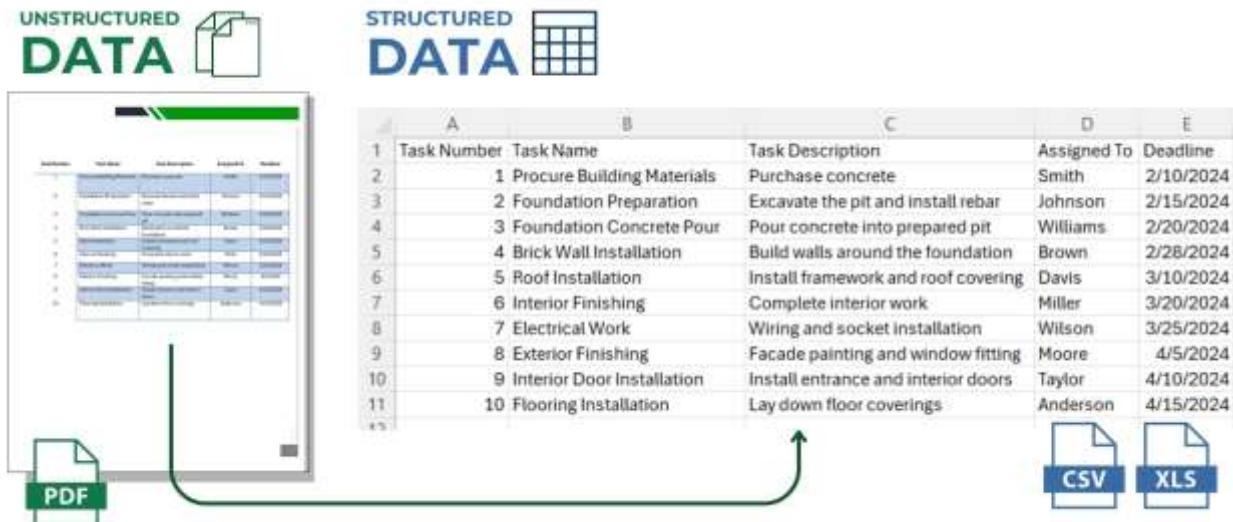
dokumentumok, a PDF, a képek és a szövegek strukturált formátummá alakítása (4.1-1. ábra) egy lépésről lépésre történő folyamat, amely a következő lépéseket foglalja magában:

- **Kivonatolás:** Ebben a lépésben egy struktúrálatlan adatokat tartalmazó forrásdokumentum vagy kép betöltése történik. Ez lehet például egy PDF -dokumentum, egy fénykép, egy rajz vagy egy vázlat.
- **Adattranszformáció (Transform):** Ezt követi a struktúrálatlan adatok strukturált formátumba történő átalakításának lépése. Ez magában foglalhatja például a szöveg felismerését és értelmezését képekről optikai karakterfelismerés (OCR) vagy más feldolgozási módszerek segítségével.
- **Az adatok betöltése és mentése (Load):** az utolsó lépés a feldolgozott adatok mentése különböző formátumokban, például CSV, XLSX, XML, JSON, a további munka számára, ahol a formátum kiválasztása az egyedi követelményektől és preferenciáktól függ.

Ez az ETL (Extract, Transform, Load) néven ismert folyamat kulcsszerepet játszik az automatizált adatfeldolgozásban, és részletesebben az "ETL és Pipeline: Extract, Transform, Load" című fejezetben tárgyaljuk. Ezután példákat fogunk megvizsgálni arra, hogy a különböző formátumú dokumentumok hogyan alakulnak át strukturált adatokká.

Példa egy PDF -dokumentum táblázattá alakítására

Az építési projektek egyik leggyakoribb feladata a PDF formátumú műszaki leírások feldolgozása. A struktúrálatlan adatok strukturált formátumba való átmenetének bemutatására tekintsünk egy gyakorlati példát: egy táblázat kivonása egy PDF-dokumentumból és átalakítása CSV- vagy Excel-formátumba (4.1-2. ábra).



Ábra 4.1-2 A PDF formátummal ellentétben a CSV és az XLSX formátum széles körben elterjedt és könnyen integrálható a különböző adatkezelési rendszerekbe.

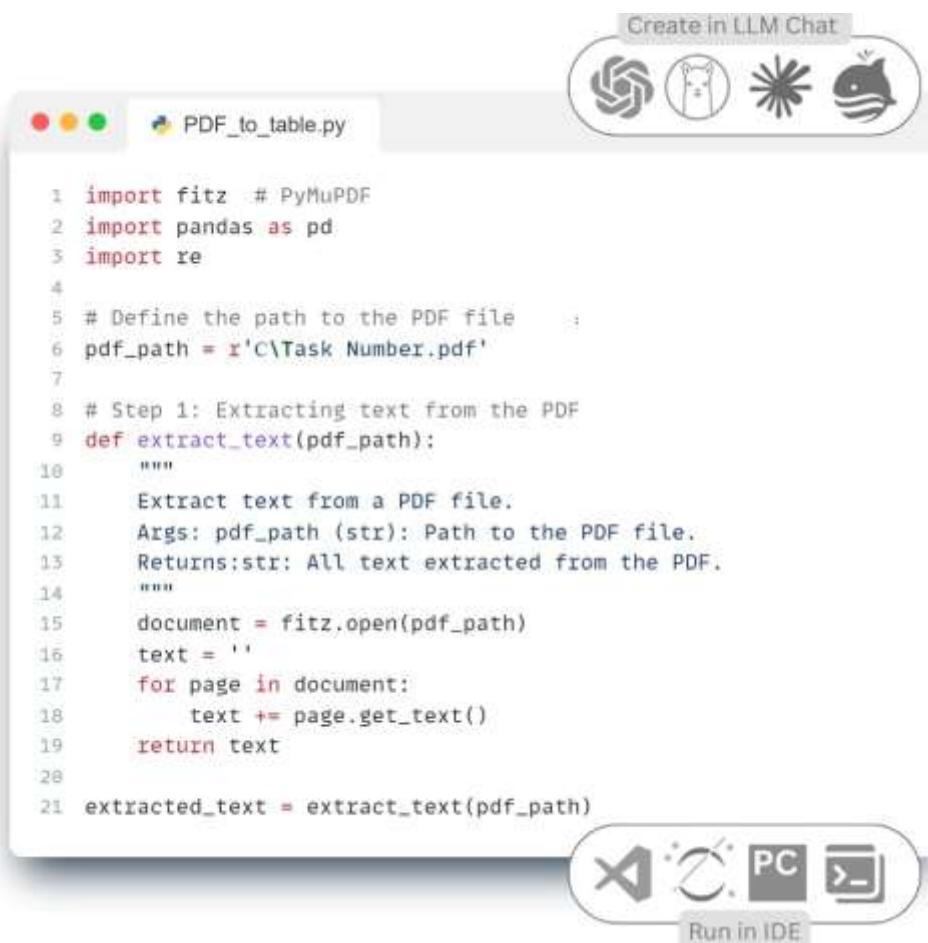
Az LLM nyelvi modellek, mint például a ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN nagyban leegyszerűsítik az adattudósok adatfeldolgozását, csökkentve a programozási nyelvek mély tanulásának szükségességét, és lehetővé téve számos feladat megoldását szöveges lekérdezésekkel.

Ezért ahelyett, hogy az interneten (általában a StackOverflow weboldalon vagy tematikus fórumokon és chateken) keresnék megoldásokat, vagy adatfeldolgozó szakemberekkel vennénk fel a kapcsolatot, a modern online vagy helyi LLM-ek lehetőségeit használhatjuk. Elég egy lekérdezést feltenni, és a modell kész kódöt ad a PDF -dokumentum táblázatos formátumba történő konvertálásához.

- ❷ Küldje el a következő szöveges kérést bármely LLM -modellnek (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vagy bármely más):

Kérem, írjon egy kódot a szöveg kivonására egy táblázatot tartalmazó PDF -fájlból. A kódnak a fájl elérési útvonalát kell argumentumként elfogadnia, és a kinyert táblázatot DataFrame-ként kell visszaadnia ↴

- Az LLM -modell válasza a legtöbb esetben Python kód formájában érkezik, mivel ezt a nyelvet széles körben használják adatfeldolgozásra, automatizálásra és különböző fájlformátumokkal való munkára:



```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11     Extract text from a PDF file.
12     Args: pdf_path (str): Path to the PDF file.
13     Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

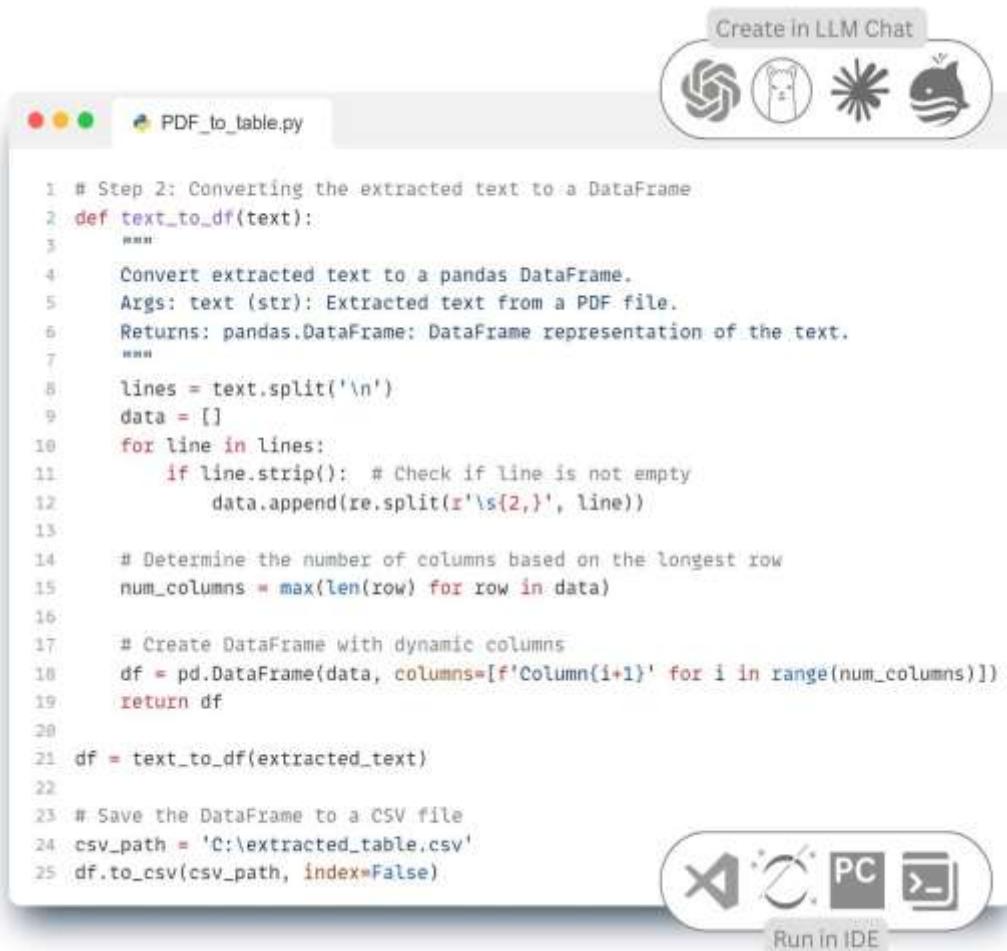
Ábra 4.1-3 Az LLM válasz a Python kód és a hozzá tartozó könyvtárak és csomagok (Pandas, Fitz) formájában szövegeket von ki egy PDF -fájlból.

Ez a kód (4.1-3. ábra) offline futtatható a fent említett népszerű IDE-k valamelyikében: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev pluginnal, Thonny, Wing IDE, IntelliJ IDEA Python pluginnal, JupyterLab vagy népszerű online eszközök: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- A "Konvertálás" lépében a népszerű Pandas könyvtárat használjuk (amelyet a "Python Pandas: egy nélkülözhetetlen eszköz az adatokkal való munkához" című fejezetben részletesen tárgyalunk), hogy a kinyert szöveget beolvassuk a DataFrame-be és elmentsük a DataFrame-et egy CSV táblázatos fájlba vagy XLXS:

Olyan kódra van szükségem, amely a PDF -fájlból származó táblázatot DataFrame -fájlba konvertálja. Továbbá kódot kell hozzáadni a DataFrame mentéséhez egy CSV fájlba. ↴

💡 LLM válasza :



The screenshot shows a code editor window titled "PDF_to_table.py". The code is as follows:

```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

At the top right of the code editor, there is a "Create in LLM Chat" button and a row of icons representing different AI models. At the bottom right, there is a "Run in IDE" button.

Ábra 4.1-4 A kivont táblázat PDF-ból DataFrame-be konvertálása és a táblázat mentése egy CSV -fájlba.

Ha a kód végrehajtása során hiba lép fel (4.1-3. ábra, 4.1-4. ábra) - pl. hiányzó könyvtárak vagy rossz fájlútvonal miatt - a hiba szövege egyszerűen bemásolható a forráskóddal együtt, és újra elküldhető az LLM -modellnek. A modell elemzi a hibaüzenetet, elmagyarázza, hogy mi a probléma, és javításokat vagy további lépéseket javasol.

Ily módon az AI LLM-mel való interakció egy teljes ciklussá vá---teszt→ visszajelzés→ korrekció - anélkül, hogy mély technikai ismeretekre lenne szükség.

Egy egyszerű szöveges lekérdezéssel az LLM chat-ben és egy tucatnyi Python sorral, amelyet helyben bármelyik IDE-ben futtathatunk, egy PDF -dokumentumot táblázatos CSV formátumúvá alakítottunk, amely a PDF dokumentummal ellentétben könnyen gépileg olvasható és gyorsan integrálható bármelyik adatkezelési rendszerbe.

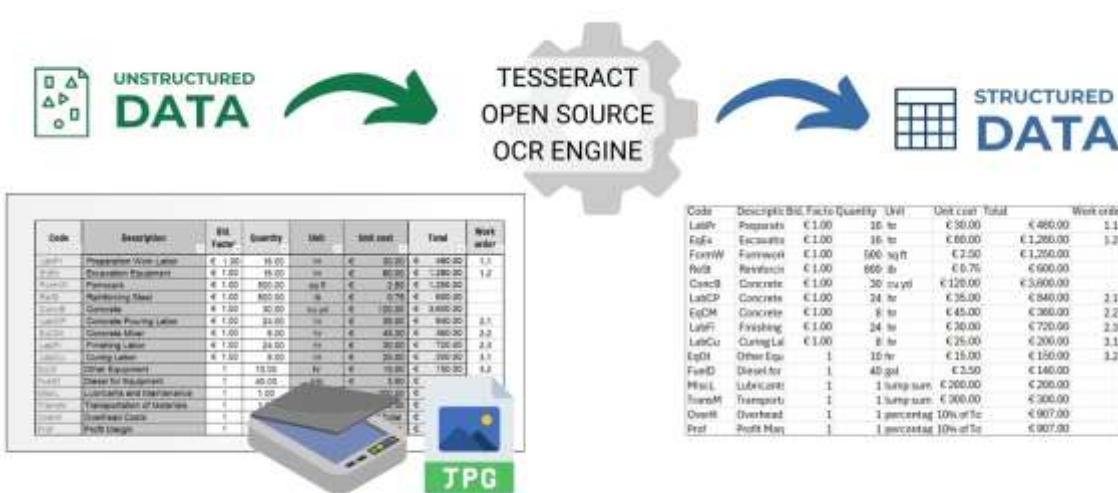
Ezt a kódot (4.1-3. ábra, 4.1-4. ábra) bármelyik LLM csevegőszobából másolva alkalmazhatjuk a szerveren lévő több tíz vagy ezer új PDF-dokumentumra, automatizálva ezzel a strukturálatlan dokumentumáradat strukturált CSV-táblázat formátumúvá alakításának folyamatát.

A PDF-dokumentumok azonban nem mindig tartalmaznak szöveget, leggyakrabban szkennelt dokumentumokról van szó, amelyeket képként kell feldolgozni. Bár a képek természetüknél fogva strukturálatlanok, a felismerő könyvtárak kifejlesztése és alkalmazása lehetővé teszi számunkra, hogy kivonjuk, feldolgozzuk és elemezzük a tartalmukat, és így teljes mértékben kihasználjuk ezeket az adatokat az üzleti folyamatainkban.

JPEG, PNG kép átalakítása strukturált formába

A képek a strukturálatlan adatok egyik leggyakoribb formája. Az építőiparban és sok más iparágban hatalmas mennyiséggű információt tárolnak szkennelt dokumentumok, vázlatok, fényképek és rajzok formájában. Az ilyen adatok értékes információkat tartalmaznak, de nem lehet őket közvetlenül feldolgozni, például Excel-táblázatban vagy adatbázisban. A képek sok összetett információt tartalmaznak, mivel tartalmuk, színük, textúrájuk változatos, és a hasznos információk kinyeréséhez speciális feldolgozásra van szükség.

A képek adatforrásként való felhasználásának nehézsége a struktúra hiánya. A képek nem közvetítenek olyan közvetlen, könnyen számszerűsíthető módon jelentést, amelyet egy számítógép azonnal megérthat vagy feldolgozhat, mint egy Excel-táblázat vagy egy adatbázis-tábla. A strukturálatlan képi adatok strukturált formába való átalakításához speciális könyvtárakat kell használni, amelyek képesek értelmezni a bennük lévő vizuális információt (4.1-5. ábra).



Ábra 4.1-5 A beolvasott dokumentumok és képek strukturált formátumba történő átalakítása speciális OCR eszközökkel lehetséges.

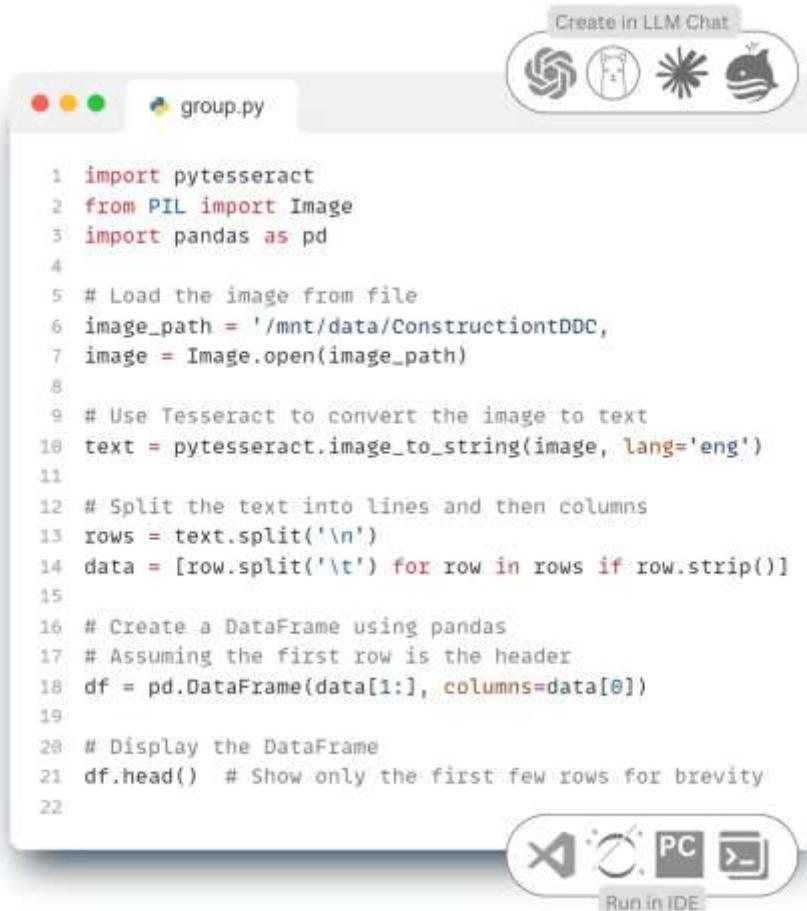
Az OCR (optikai karakterfelismerés) technológiát a szöveg képekből történő kinyerésére használják. Lehetővé teszi a betűk és számok felismerését a dokumentumszkennelésekben, fényképekben és PDF -fájlokban, szerkeszthető és géppel olvasható szöveggé alakítva azokat. Az OCR technológiákat régóta használják a dokumentumautomatizálásban, és ma már könnyen integrálhatók bármely üzleti folyamatba és Python -alkalmazásba. Az egyik legnépszerűbb OCR-eszköz a Tesseract, egy nyílt forráskódú, amelyet eredetileg a HP™ fejlesztett ki, és ma már a Google™ is támogat. Több mint 100 nyelvet támogat, és nagy felismerési pontossággal rendelkezik.

Kérdezzük meg az LLM chat-et, hogy írjon mintakódot az adatok strukturált módon történő lekérdezésére egy beolvasott vagy lefényképezett táblázatból.

- ✉ Küldjön szöveges kérést az LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vagy bármely más):

Írj kódot egy táblázatot tartalmazó JPEG kép DataFrame táblázattá történő konvertálásához ↴

- Az LLM válasza a legtöbb esetben a Pytesseract könyvtár használatát javasolja a képeken található szövegek felismeréséhez :



```

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC'
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22

```

Ábra 4.1-6 Képek vagy fényképek táblázatából kinyert szöveg átalakítása strukturált táblázatos ábrázolássá.

Ebben a példában - az LLM -ból származó kód (4.1-6. ábra) a pytesseract könyvtárat (Tesseract for Python) használja a kép OCR (optikai karakterfelismerés) segítségével történő szöveggé alakítására, és a Pandas könyvtárat a szöveg strukturált formába, azaz egy DataFrame -be való átalakítására.

Az átalakítási folyamat általában előfeldolgozást tartalmaz a képminőség javítása érdekében, majd különböző algoritmusokat alkalmaznak a minta detektálására, a jellemzők kinyerésére vagy a tárgyfelismerésre. Ennek eredményeként a strukturálatlan vizuális információ strukturált adatokká alakul át.

Bár a PDF és a képek a strukturálatlan információk kulcsfontosságú forrásai, a igazi bajnoka a mennyiséget tekintve az e-mailben, csevegésekben, megbeszéléseken és üzenetküldőkben keletkező szöveg. Ezek az adatok nem csak nagyszámúak, hanem szétszórtak, informálisak és rendkívül rosszul strukturáltak.

Szöveges adatok strukturált formába történő átalakítása

A táblázatokat tartalmazó PDF-dokumentumok (4.1-2. ábra) és a táblázatos űrlapok szkennelt változatai (4.1-5. ábra) mellett a projektdokumentációban szereplő információk jelentős része szöveges formában kerül bemutatásra. Ez lehet szöveges dokumentumokban szereplő összefüggő mondatok és a rajzokon és sémaikon elszórtan található töredékes feljegyzések egyaránt. A modern adatfeldolgozási körülmények között az egyik leggyakoribb feladat az ilyen szövegek az elemzésre, megjelenítésre és döntéshozatalra alkalmas strukturált formátumba történő átalakítása.

E folyamat központi eleme a taxonómia, egy olyan osztályozási rendszer, amely lehetővé teszi az információk kategóriákba és alkategóriákba rendezését közös jellemzők alapján.

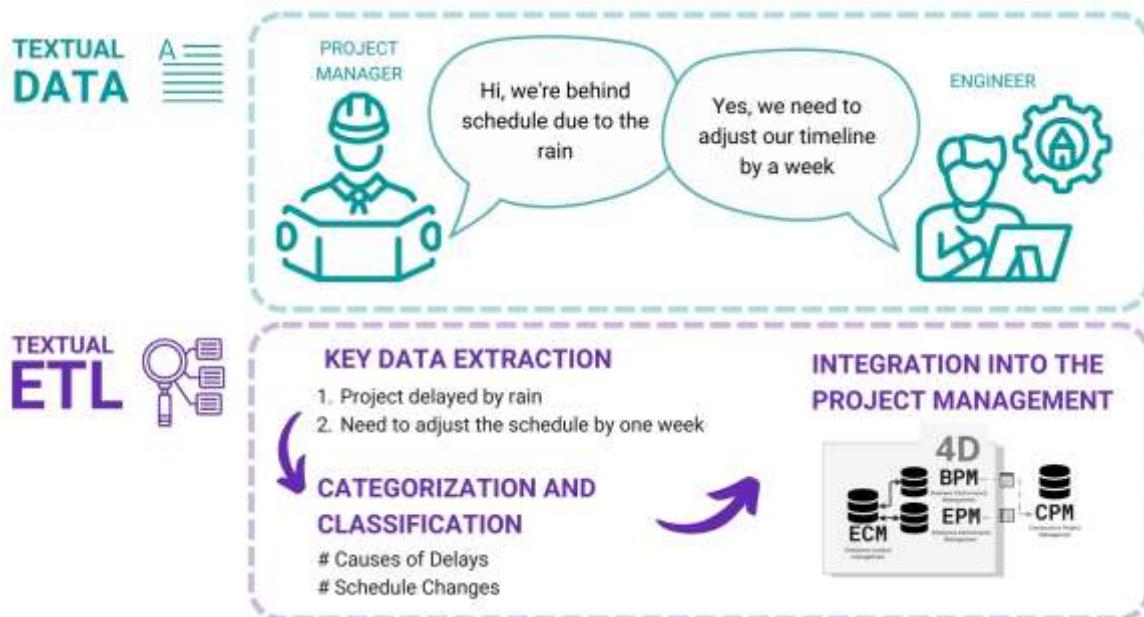
A taxonómia egy hierarchikus osztályozási struktúra, amelyet az objektumok csoportosítására és rendszerezésére használnak. A szövegfeldolgozásban az elemek szemantikai kategóriákba történő szisztematikus kategorizálásának alapjául szolgál, ezáltal egyszerűsítve az elemzést és javítva az adatfeldolgozás minőségét.

A taxonómia létrehozását az entitások kinyerésének, kategorizálásának és kontextualizálásának lépései kísérlik. A szöveges adatokból történő információszerzés folyamatának modellezéséhez a következő lépések hasonlóak azokhoz, amelyeket már alkalmaztunk a PDF-dokumentumokból származó adatok strukturálására:

- Extract): szöveges adatokat kell elemeznie a projekt ütemtervében bekövetkezett késésekkel és változásokkal kapcsolatos információk kinyerése érdekében.
- Kategorizálás és osztályozás (Transofrm): a kapott információk kategorizálása, pl. a késések és menetrendi változások okai.
- Integráció (Load): a végén strukturált adatokat készítünk elő a külső adatkezelési rendszerekbe történő integrációhoz.

Vegyük egy helyzetet: egy projektmenedzser és egy mérnök között párbeszéd folyik az ütemtervvel kapcsolatos problémákról. Célunk, hogy a kulcsfontosságú elemeket (késedelem okai, ütemterv-kiigazítások) kiemeljük és strukturáltan bemutassuk (4.1-7. ábra).

Végezzük el a kivonatolást a várt kulcsszavak alapján, hozunk létre egy DataFrame adatkivonatot az adatok kivonatolásának szimulálására, majd az átalakítás után egy új DataFrame táblát, amely a dátum, az esemény (pl. a késés oka) és a művelet (pl. menetrendváltozás) oszlopait tartalmazza.



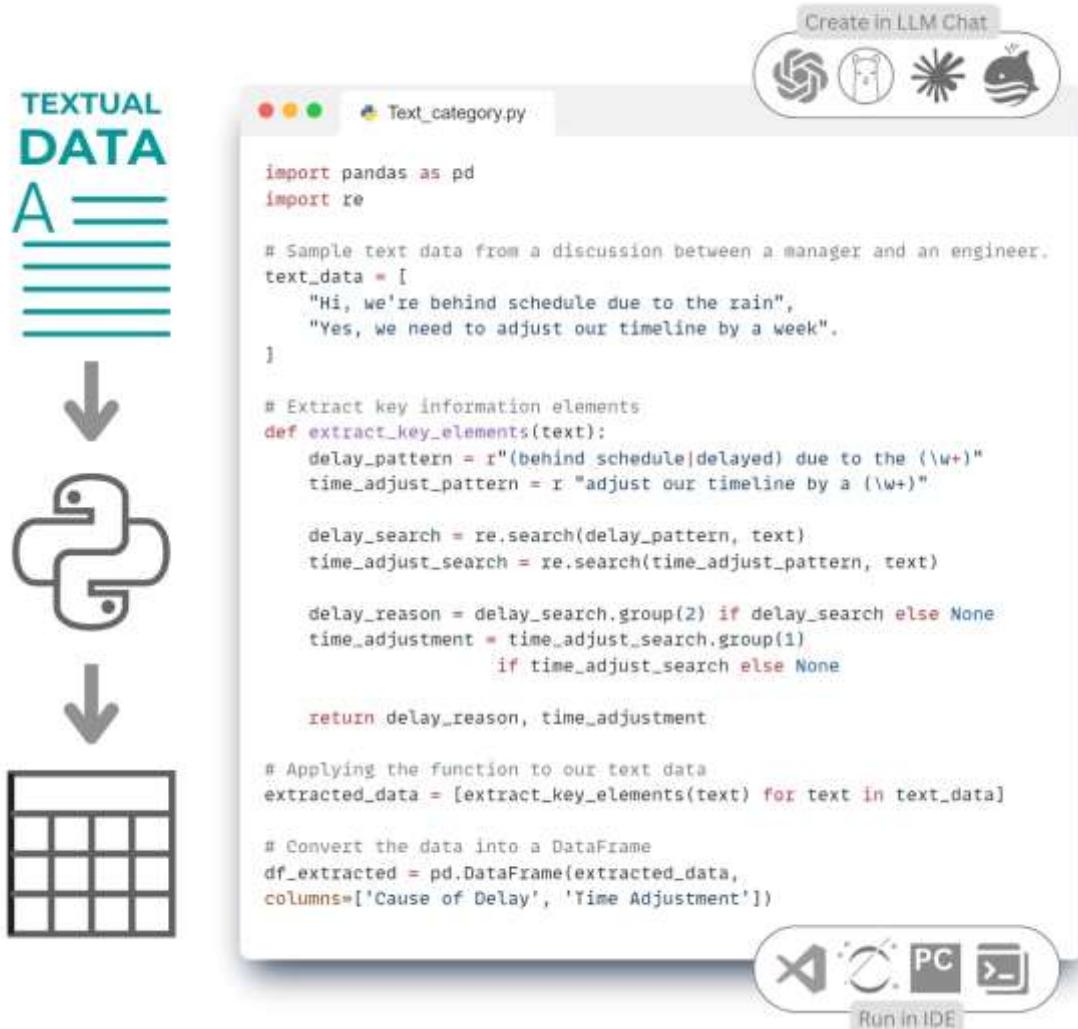
4.1-7. ábra A szövegben szereplő kulcsfontosságú információk kiemelése az ütemtervezés kiigazításának és a változások projektirányítási rendszerbe való integrálásának szükségesből.

Íme a kód a probléma megoldásához egy szöveges lekérdezéssel az egyik nyelvi modellben, mint az előző példákban.

❷ Küldjön szöveges kérést bármelyik LLM csevegőszobába:

Beszélgetést folytatok egy menedzserrel: "Helló, az eső miatt késésben vagyunk", és egy mérnökkel: "Igen, egy héttel módosítani kell a határidőt". Szükségem van egy olyan szkriptre, amely elemzi a jövőbeli hasonló szöveges párbeszédeket, kivonja belőlük a késések okait és a szükséges határidő-kiigazításokat, majd ezekből a adatokból létrehoz egy DataFrame-et. A DataFrame-et ezután el kell menteni egy CSV -fájlba. ↵

- Az LLM válasza jellemzően a Python -kódot tartalmazza, amely a reguláris kifejezéseket (re - Regex) és a Pandas (pd) könyvtárat használja:

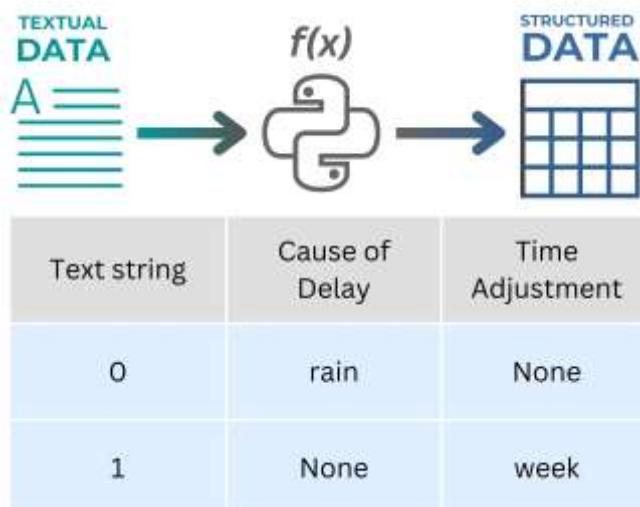


4.1-8 ábra A határidők módosításának szükségességére vonatkozó kulcsfontosságú információk kiemelése a szövegből egy táblázatban.

Ebben a példában (4.1-7. ábra) egy projektmenedzser és egy mérnök közötti levelezést tartalmazó szöveges adatokat elemeznek, hogy azonosításuk és kivonják azokat a konkrét információkat, amelyek befolyásolhatják a hasonló párbeszédet folytató jövőbeli projektek irányítását. szabályos kifejezések használatával (a szabályos kifejezésekkel bővebben a "Strukturált követelmények és RegEx szabályos kifejezések" című fejezetben lesz szó) a projekt késedelmének okai és az ütemterv szükséges kiigazításai mintákon keresztül azonosíthatók. A példában leírt függvény a minták alapján kivonja a karakterláncokból a késés okát vagy az időbeli kiigazítást: a "because of" utáni szót választja ki a késés okaként, illetve a "by" utáni szót az időbeli kiigazításként.

Ha egy sor időjárás miatti késést említi, akkor az "eső" lesz az ok; ha egy sor egy bizonyos időszakra

vonatkozó menetrendi kiigazítást említ, akkor ez az időszak lesz kivonva időbeli kiigazításként (4.1-9. ábra). Ha egy sorban nincs ilyen szó, a megfelelő attribútum-oszlopban a "Nincs" értéket kapja.



Ábra 4.1-9 A kód végrehajtása után a DataFrame formában kapott összefoglaló táblázat tartalmazza a késések meglétéről és a szükséges időbeli kiigazításokról szóló információkat.

A feltételek strukturálása és paraméterezése a szövegből (párbeszéd, levél, dokumentum) lehetővé teszi az építkezésben bekövetkező késések gyors kiküszöbölését: például a munkások hiánya rossz időjárás esetén befolyásolhatja a munka ütemét, így a vállalatok, ismerve a késés paramétereit az építkezésen lévő művezető és a projektvezető közötti párbeszédekből (4.1-9. ábra) - előre - megerősíthetik a személyzetet kedvezőtlen előrejelzés esetén.

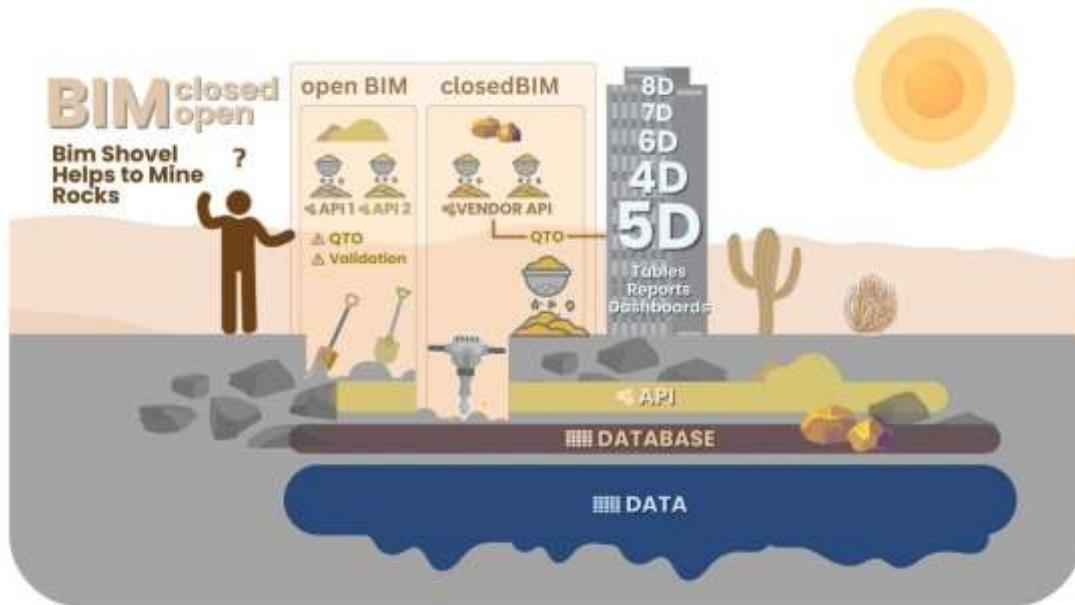
A dokumentumok és képek strukturált formátumba történő átalakítása viszonylag egyszerű, nyílt és ingyenes kategorizáláson alapuló eszközökkel valósítható meg.

Az elemek kategorizálása szintén kulcsfontosságú része a projektadatokkal való munkának, különösen a CAD-szoftverek használatával összefüggésben (BIM).

A CAD-adatok (BIM) strukturált formába történő konvertálása

A CAD-adatok strukturálása és kategorizálása (BIM) összetettebb feladat, mivel a CAD (BIM) adatbázisokban tárolt adatok szinte mindenkorán zárt vagy összetett parametrikus formátumban vannak, és gyakran egyszerre kombinálják a geometriai adatelemeket (félig strukturált) és a metainformációs elemeket (félig strukturált vagy strukturált adatok).

A CAD (BIM) rendszerek natív adatformátumai általában védettek és közvetlen használatra hozzáférhetetlenek, kivéve, ha a fejlesztő saját maga speciális vagy API - interfészket használ (4.1-10. ábra). Az adatoknak ez az elszigeteltsége zárt tárolási silókat képez, amelyek korlátozzák a szabad információcserét és gátolják a vállalaton belüli végponttól végpontig tartó digitális folyamatok kialakítását.



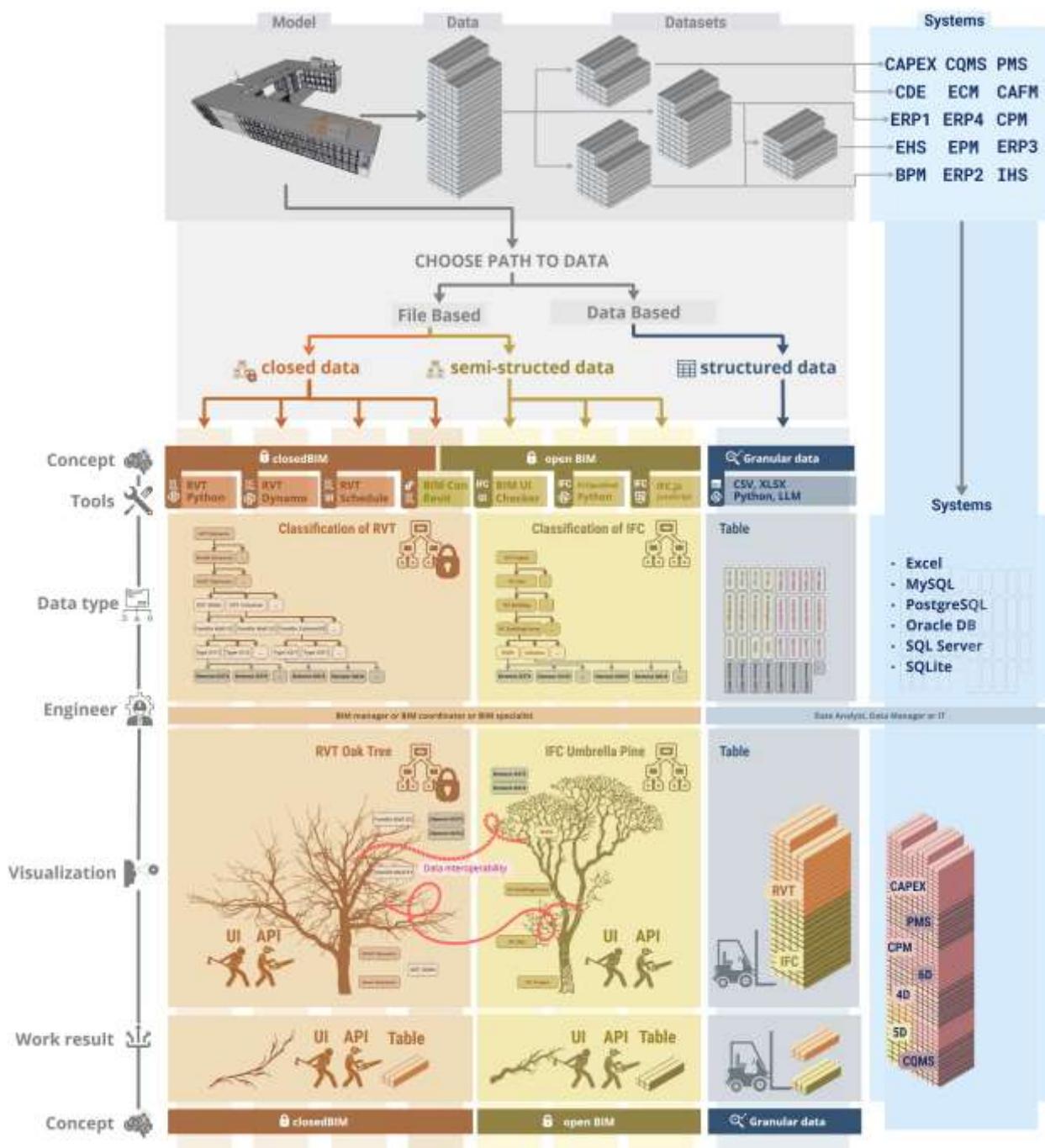
4.1-10. ábra A CAD szakemberek (BIM) az API -kapcsolatokon vagy a gyártók eszközein keresztül férhetnek hozzá a natív adatokhoz.

A speciális CAD (BIM) formátumokban a projektelemek jellemzőire és attribútumaira vonatkozó információkat egy hierarchikus osztályozási rendszerben gyűjti össze, ahol a megfelelő tulajdonságokkal rendelkező entitások a gyümölcsfához hasonlóan a adatosztályozási ágak legfrissebb csomópontjaiban helyezkednek el (4.1-11. ábra).

Az adatok kinyerése az ilyen hierarchiából kétféleképpen lehetséges: vagy manuálisan, az egyes csomópontokra kattintva, mintha egy fát dolgoznánk fel, a kategóriák és típusok kiválasztott ágait fejszével levágva. Alternatív megoldásként az alkalmazásprogramozási interfészök (API-k) használata hatékonyabb, automatizált megközelítést biztosít az adatok kinyerésére és csoportosítására, és végül más rendszerekben való felhasználásra alkalmas strukturált táblázattá alakítására.

A CAD (BIM) projektekből strukturált adattáblák kinyerésére különböző eszközök, például Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC vagy nyílt forráskódú megoldások, például IfcOpSh vagy IFCjs az IFC formátumhoz, használhatók.

A modern adatexport és -konverziós eszközök lehetővé teszik az adatfeldolgozás és -előkészítés egyszerűsítését azáltal, hogy a CAD-modellek tartalmát két fő összetevőre osztják: geometriainformáció és attribútumadatok (4.1-13. ábra) - a tervezési elemek tulajdonságait leíró metainformációk (3.1-16. ábra). Ez a két adatréteg egyedi azonosítókon keresztül marad összekapcsolva, amelyeknek köszönhetően minden egyes geometrialeírást tartalmazó elemet (paraméterek vagy poligonok segítségével) pontosan le lehet rendelni az attribútumokhoz: név, anyag, készültségi fok, költség stb. révén. Ez a megközelítés biztosítja a modell integritását, és lehetővé teszi az adatok rugalmas felhasználását mind a vizualizációhoz (geometriai modelladatok), mind az elemzési vagy kezelési feladatokhoz (strukturált vagy lazán strukturált), a kétféle adattípussal külön-külön vagy párhuzamosan dolgozva.



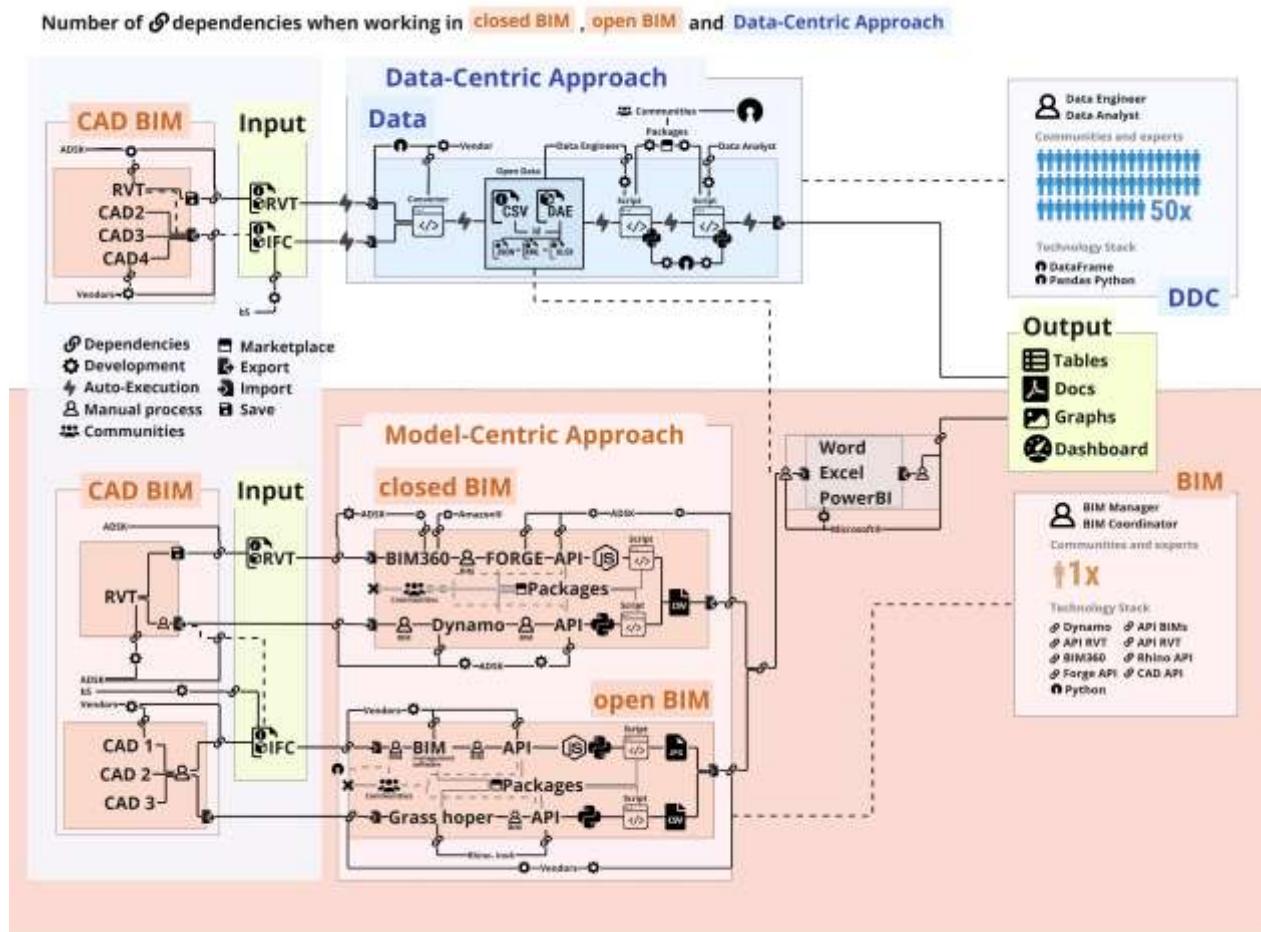
Ábra 4.1-11 A CAD-adatbázisokból (BIM) származó információk nézete a felhasználó számára osztályozási fák formájában jelenik meg.

A reverse engineering technológiák fejlődésével és a CAD-adatok konvertálására szolgáló SDK-k megjelenésével (Software Development Kit) - a zárt CAD-szoftverformátumok (BIM) adatainak elérhetősége és konvertálása sokkal egyszerűbbé vált. Ma már lehetséges a zárt formátumokból származó adatok legalisan és biztonságosan átalakítása olyan univerzális formátumokba, amelyek alkalmasak az elemzésre és más rendszerekben való felhasználásra. Az első reverse engineering eszközök ("Open DWG") történetét

és a CAD-gyártók formátumai feletti dominanciáért folytatott küzelmet a "Strukturált adatok: a digitális átalakulás alapja" című fejezetben tárgyaltuk.

A Reverse Engineering eszközök lehetővé teszik az adatok törvénies visszakeresését zárt, védett formátumokból, a vegyes CAD (BIM) formátumból származó információkat a felhasználó által igényelt adattípusokra és formátumokra bontják, megkönnyítve azok feldolgozását és elemzését.

a CAD-adatbázisokból származó információkhoz való közvetlen hozzáférés és a reverse engineering segítségével hozzáférhetővé teszi az információkat, lehetővé téve a nyílt adatok és nyílt eszközök használatát, valamint az adatok szabványos eszközökkel történő elemzését, a jelentések készítését, a vizualizációt és az egyéb digitális rendszerekkel való integrációt (4.1-12. ábra).



4.1-12 A CAD-adatokhoz való közvetlen hozzáférés minimalizálja a szoftverplatformktól való függőséget, és az adatközpontú megközelítés felé mozdul el.

1996 óta a DWG formátum, 2008 óta a DGN formátum és 2018 óta az RVT esetében a kezdetben zárt CAD adatformátumok bármely más formátumba, beleértve a strukturált formátumokat is, kényelmesen és hatékonyan konvertálhatók a reverse engineering eszközök segítségével (4.1-13. ábra). Ma már a világ szinte valamennyi jelentős CAD (BIM) és nagy mérnöki vállalat használja a SDK-keket - reverse engineering

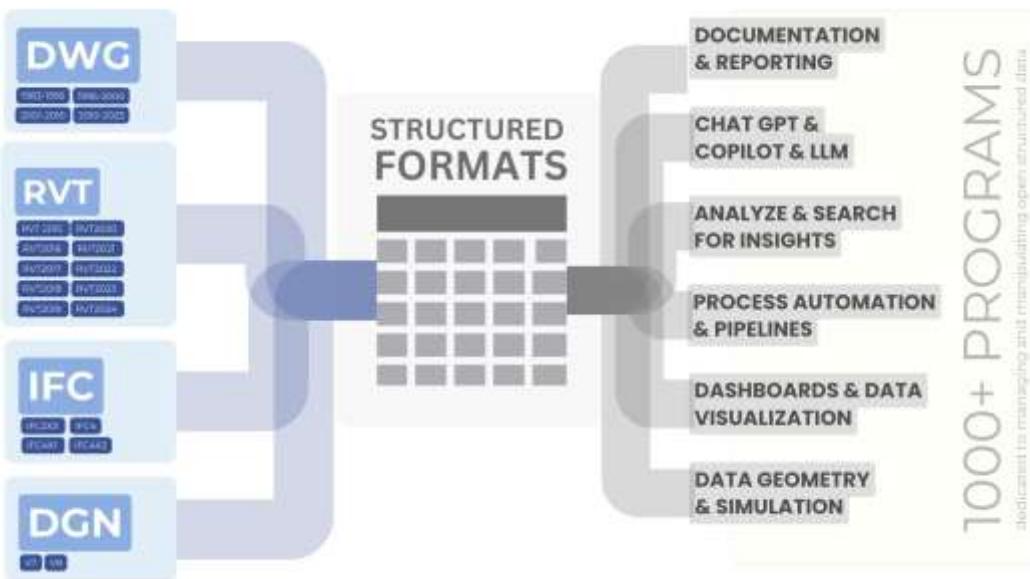
eszközöket - a zárt CAD (BIM) szállítói formátumokból történő adatkivonáshoz [92].



Ábra 4.1-13 A reverse engineering eszközök használata lehetővé teszi a CAD (BIM) programadatbázisok átalakítását bármilyen kényelmes adatmodellé.

Az adatok zárt, védett formátumokból nyílt formátumokba történő átalakítása és a vegyes CAD (BIM) formátumok szétválasztása geometriai és metainformációs attribútumadatokra egyszerűsíti a velük való munkavégzés folyamatát, elérhetővé téve azokat az elemzés, manipuláció és más rendszerekkel való integráció számára (4.1-14. ábra).

A CAD-adatokkal való mai munkánk során (BIM) elérkeztünk arra a pontra, amikor már nem kell engedélyt kérnünk a CAD (BIM) gyártótól ahhoz, hogy hozzáférjünk a CAD-formátumokból származó információkhoz.



Ábra 4.1-14 A modern SDK-eszközök lehetővé teszik az adatok legális konvertálását a saját CAD-adatbázisformátumokból (BIM).

A CAD tervezési adatok feldolgozásának jelenlegi trendjeit továbbra is a legfontosabb piaci szereplők - CAD - alakítják, akik azon dolgoznak, hogy megerősítésük pozíciójukat az adatok világában, és új formátumokat és koncepciókat hozzanak létre.

A CAD megoldások szállítói a strukturált adatok felé mozdulnak el.

2024-től kezdve a tervezési és építési ágazat jelentős technológiai váltáson megy keresztül az adatok felhasználása és feldolgozása terén. A tervezési adatokhoz való szabad hozzáférés helyett a CAD - rendszerek szállítói a következő új koncepciók népszerűsítésére összpontosítanak. Az olyan megközelítések, mint a BIM (2002-ben jött létre) és a nyílt BIM (2012-ben jött létre) fokozatosan átadják helyüket a modern technológiai megoldásoknak, amelyeket a CAD-gyártók kezdenek népszerűsíteni [93]:

- A "granuláris" adatok használata felé való elmozdulás, amely lehetővé teszi a hatékony információkezelést és az adatelemzés felé való elmozdulást.
- Az USD formátum megjelenése és az Entity-component-system megközelítés (ECS) bevezetése a rugalmas adatszervezés érdekében.
- A mesterséges intelligencia aktív használata az adatfeldolgozásban, a folyamatok automatizálásában és az adatelemzésben.
- Az interoperabilitás fejlesztése - jobb interakció a különböző programok, rendszerek és adatbázisok között.

E szempontok mindegyikét részletesebben a "CAD és BIM: marketing, valóság és a tervezési adatok jövője az építőiparban" című könyv hatodik része tárgyalja. Ebben a fejezetben csak röviden vázoljuk fel a változások általános vektorát: a nagy CAD-gyártók most arra törekszenek, hogy újragondolják a tervezési információk strukturálásának módját. Az egyik legfontosabb váltás a klasszikus fájlalapú tárolási modell elhagyása egy granuláris, analitika-orientált adatarchitektúra javára, amely folyamatos hozzáférést biztosít az egyes modellkomponensekhez [93].

A folyamat lényege, hogy az iparág fokozatosan kivonja a nehézkes, speciális és parametrikus formátumokat, amelyek geometriai magokat igényelnek, az univerzálisabb, géppel olvasható és rugalmasabb megoldások javára.

A változás egyik ilyen mozgatórugója az USD (Universal Scene Description) formátum, amelyet eredetileg a számítógépes grafikai iparban fejlesztettek ki, de a szimulációk és vizualizációk számára kifejlesztett NVIDIA Omniverse (és Isaac Sim) platformnak köszönhetően már a mérnöki alkalmazásokban is elismert [93]. A parametrikus IFC -től eltérően az USD egyszerűbb struktúrát kínál, és lehetővé teszi a geometria és az objektum tulajdonságainak JSON formátumban történő leírását (4.1-15. ábra), ami megkönnyíti az információfeldolgozást és felgyorsítja a digitális folyamatokba való integrálást. Az új formátum lehetővé teszi a geometria (a BREP -NURBS mellett - további részletek a könyv 6. részében) MESH-poligonok formájában történő tárolását, valamint az objektumtulajdonságok JSON-ban történő tárolását, ami kényelmesebbé teszi az automatizált folyamatokat és a felhőalapú ökoszisztemákban történő munkát [94].

Néhány CAD- és ERP-szállító már használ hasonló formátumokat (pl. NWD, SVF, CP2, CPIXML), de ezek többsége továbbra is zárt és külső használatra nem elérhető, ami korlátozza az adatintegráció és újrafelhasználás lehetőségeit. Ebben az összefüggésben az USD ugyanazt a szerepet játszhatja, mint annak idején a DXF - nyílt alternatívát jelenthet az olyan védett formátumokkal szemben, mint a DWG.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in bS (IAI) since the introduction
Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

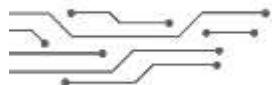
Ábra 4.1-15 USD formátum a CAD gyártók kísérlete arra, hogy kielégítsék az interoperabilitás és a tervezési adatok geometriai magoktól való függetlensége iránti igényt.

A főbb fejlesztők áttérése a nyílt és egyszerűsített USD, GLTF, OBJ, XML (zárt NWD, CP2, SVF, SVF2, CPIXML) és hasonló formátumokra (3.1-17. ábra) tükrözi az adatok egyszerűsítése és a fokozott hozzáférhetőség iránti globális tendenciát és ipari igényt. Az elkövetkező években várhatóan fokozatosan eltávolodunk a geometriai magoktól függő, összetett paraméteres szabványuktól és formátumuktól a könnyebb és strukturáltabb megoldások javára. Ez az átmenet fel fogja gyorsítani az építőipar digitalizációját, megkönnyíti a folyamatok automatizálását és egyszerűsíti az adatcserét.

Annak ellenére, hogy a CAD -szolgáltatók stratégiai tervei az új nyílt formátumok népszerűsítésére irányulnak, az építőipari szakemberek a CAD (BIM) eszközök nélkül is teljes hozzáférést kaphatnak a zárt CAD rendszerek adataihoz, a fordított mérnöki eszközök használatával.

Mindezek a trendek elkerülhetetlenül a terjedelmes, monolitikus 3D modellek ről az univerzális, strukturált adatokra és a más iparágakban már régóta bevált formátumok használatára való áttéréshez vezetnek. Amint a projektcsapatok a CAD-modelleket nem csupán vizuális objektumként vagy fájlkészletként, hanem tudást és információt tartalmazó adatbázisokként kezdi kezelni, a tervezés és a menedzsment megközelítése drámaian megváltozik.

Miután a csapatok megtanulták, hogyan lehet strukturált adatokat kinyerni a dokumentumokból, szövegekből, rajzokból és CAD-modellekből, és hozzáférnek az adatbázisokhoz, a következő kulcsfontosságú lépés az adatmodellezés és a minőségbiztosítás. Ez a lépés nagymértékben meghatározza a feldolgozás sebességét és az információk átalakítását, amelyeket végül a konkrét alkalmazási feladatokban történő döntéshozatalhoz használnak fel.



FEJEZET 4.2.

OSZTÁLYOZÁS ÉS INTEGRÁCIÓ: AZ ÉPÍTÉSI ADATOK KÖZÖS NYELVE

A döntéshozatal sebessége az adatok minőségétől függ

A mai tervezési adatarchitektúra alapvető változásokon megy keresztül. Az iparág a terjedelmes, elszigetelt modellektől és a zárt formátumuktól a rugalmasabb, gépileg olvasható struktúrák felé mozdul el, amelyek az analitikára, az integrációra és a folyamatok automatizálására összpontosítanak. Az új formátumokra való áttérés önmagában azonban nem garantálja a hatékonyságot - maga az adatok minősége elkerülhetetlenül a figyelem középpontjába kerül.

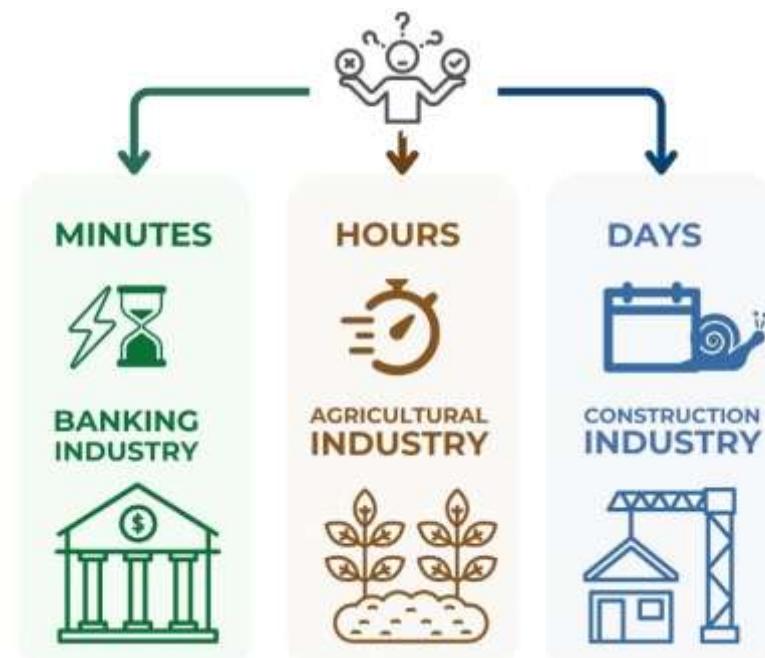
E könyv lapjain sokat beszélünk formátumokról, rendszerekről és folyamatokról. De mindenek az erőfeszítések értelmetlenek egy kulcsfontosságú elem nélkül: megbízható adatok nélkül. Az adatminőség a digitalizáció sarokköve, amelyre a következő részek során mindenkor visszatérünk.

A modern építőipari vállalatok - különösen a nagyvállalatok - több tucat, néha több ezer különböző rendszert és adatbázist használnak (4.2-1. ábra). Ezeket a rendszereket nemcsak rendszeresen új információkkal kell feltölteni, hanem hatékonyan kell együttműködniük egymással. A beérkező információk feldolgozásának eredményeként keletkező összes új adat beépül ezekbe a környezetekbe, és konkrét üzleti feladatok megoldását szolgálja.

És míg a múltban a konkrét üzleti feladatokkal kapcsolatos döntéseket a felsővezetők - az úgynevezett HiPPO-k (2.1-9. ábra) - hozták meg a tapasztalat és az intuíció alapján, addig ma, az információ mennyiségének erőteljes növekedésével ez a megközelítés ellentmondásosnak tűnik. Az automatizált analitika, amely valós idejű adatokkal dolgozik, lép a helyébe.

A "hagyományos-manuális" vezetői szintű üzleti folyamatokról szóló megbeszélések az operatív analitika irányába tolódnak el, amely gyors válaszokat igényel az üzleti kérdésekre.

Az a korszak, amikor a könyvelők, művezetők és becslők napokon és heteken át kézzel készített jelentéseket, összefoglaló táblázatokat és projektadatokat mutattak be, már a múlté. Ma a döntéshozatal gyorsasága és időszerűsége a versenyelőny kulcsfontosságú tényezőjévé válik.



Ábra 4.2-1 Az építőiparban napokig tart a számítások és a döntések meghozatala, ellentétben más iparágakkal, ahol ez órák vagy percek alatt történik.

Az építőipar és a digitálisan fejlettebb iparágak közötti fő különbség (4.2-1. ábra) az adatminőség és a szabványosítás alacsony szintje. Az információ előállításának, továbbításának és feldolgozásának elavult megközelítései lassítják a folyamatokat és káoszt okoznak. Az egységes adatminőségi szabványok hiánya akadályozza a végponttól végpontig tartó automatizálás megvalósítását

Az egyik fő kihívás továbbra is a bemeneti adatok gyenge minősége, valamint az előkészítésükre és hitelesítésükre szolgáló formalizált folyamatok hiánya. Megbízható és konzisztens adatok nélkül nem lehetséges a rendszerek közötti hatékony integráció. Ez a projekt életciklusának minden szakaszában késésekhez, hibákhoz és megnövekedett költségekhez vezet.

A könyv következő részeiben részletesen megvizsgáljuk, hogyan javíthatja az adatminőséget, hogyan szabványosíthatja a folyamatokat, és hogyan rövidítheti le az információtól a minőségi, hitelesített és konzisztens adatokig vezető utat.

Adatszabványosítás és -integráció

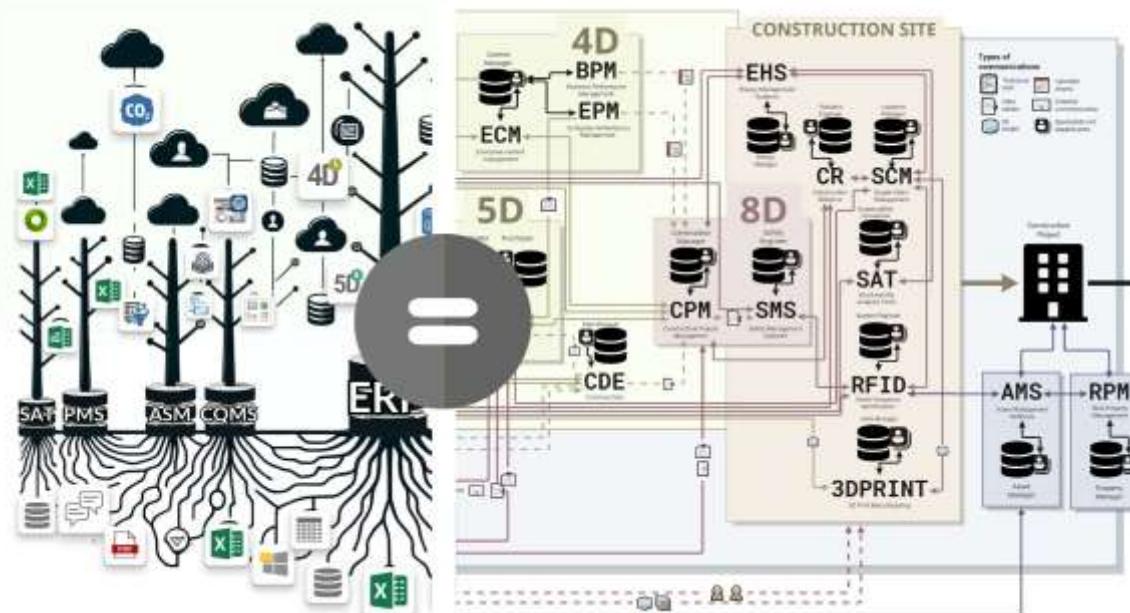
A hatékony adatkezeléshez egyértelmű szabványosítási stratégiára van szükség. Csak az adatszerkezetre és -minőségre vonatkozó egyértelmű követelményekkel lehet automatizálni az adatok érvényesítését, csökkenteni a manuális műveleteket, és felgyorsítani a megalapozott döntéshozatalt a projekt minden szakaszában.

A napi gyakorlatban egy építőipari vállalatnak naponta több száz fájlt kell feldolgozna: e-maileket, PDF - dokumentumokat, CAD tervezési fájlokat, IOT-érzékelőkből származó adatokat, amelyeket integrálni kell a

vállalat üzleti folyamataiba.

A vállalat adatbázisokból és eszközökből álló ökoszisztemájának erdejének (4.2-2. ábra) meg kell tanulnia a beérkező több formátumú adatokból tápanyagot nyerni ahhoz, hogy a vállalat által kívánt eredményeket érje el.

Az adatáramlás hatékony kezeléséhez nem feltétlenül kell menedzserek egész sereget alkalmazni, hanem először is szigorú követelményeket és szabványokat kell kidolgozni az adatokra vonatkozóan, és megfelelő eszközöket kell használni az adatok automatikus érvényesítésére, harmonizálására és feldolgozására.



Ábra: 4.2-2 Az egészséges vállalati ökoszisztemá biztosítása megköveteli a rendszerek minőségi és időben történő erőforrás-ellátását.

Az adatérvényesítési és harmonizációs folyamat automatizálása érdekében (a későbbi automatikus integráció érdekében) az egyes konkrét rendszerek minimálisan szükséges adatkövetelményeinek leírásával kell kezdeni. Ezek a követelmények határozzák meg:

- Pontosan mit kell beszereznie?
- Milyen formában (szerkezet, formátum)?
- Milyen attribútumok kötelezők?
- Milyen pontossági és teljességi tűréshatárok elfogadhatók?

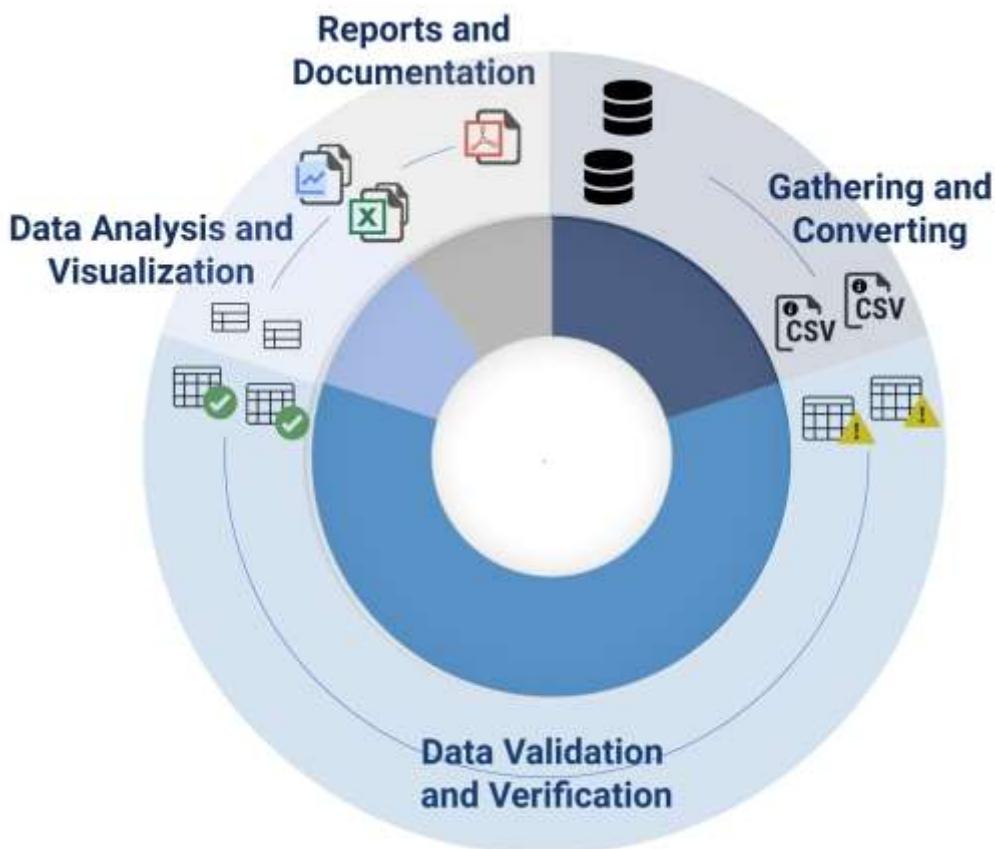
Az adatkövetelmények a kapott és feldolgozott információk minőségére, szerkezetére és teljességére vonatkozó kritériumokat írják le. Például a PDF -dokumentumokban található szövegek esetében fontos, hogy azok pontosan, az ipari szabványoknak megfelelően legyenek formázva (7.2-14. ábra - 7.2-16. ábra). A CAD -modellekben szereplő objektumoknak megfelelő attribútumokkal (méretek, kódok, osztályozókra

mutató hivatkozások) kell rendelkezniük (7.3-9. ábra, 7.3-10. ábra). A szerződések szkennelésénél pedig fontosak az egyértelmű dátumok, valamint az összeg és a kulcsfontosságú feltételek automatikus kinyerésének képessége (4.1-7. ábra - 4.1-10. ábra).

Az adatszolgáltatási követelmények megfogalmazása és azok megfelelőségének automatikus ellenőrzése az egyik legidőigényesebb, de kritikus lépés. Ez a legidőigényesebb lépés az üzleti folyamatokban.

Amint azt e könyv 3. részében említettük, az üzleti intelligencia (BI) szakemberek idejének 50%-90%-át inkább az adatok előkészítésével, mint elemzésével töltik (3.2-5. ábra). Ez a folyamat magában foglalja az adatgyűjtést, az ellenőrzést, a validálást, a harmonizálást és a strukturálást.

Egy 2016-os felmérés [95] szerint az adattudósok a legkülönbözőbb, széles spektrumú területeken azt állították, hogy munkaidejük nagy részét (körülbelül 80%-át) azzal töltik, amit a legkevésbé szeretnek csinálni (4.2-3. ábra): a meglévő adathalmazok összegyűjtésével és rendszerezésével (egységesítésével, strukturálásával). Így idejük kevesebb mint 20 százaléka marad kreatív feladatokra, például olyan minták és szabályszerűségek megtalálására, amelyek új felismerésekhez és felfedezésekhez vezetnek.



4.2-3 ellenőrzése és az adatminőség biztosítása a legköltségesebb, legidőigényesebb és legösszetettebb lépés az adatok más rendszerekbe történő integrálásra való előkészítése során.

A sikeres adatkezelés egy építőipari vállalatnál átfogó megközelítést igényel, amely magában foglalja a feladatok paraméterezését, az adatminőségi követelmények megfogalmazását és az automatizált validálásukhoz megfelelő eszközök használatát.

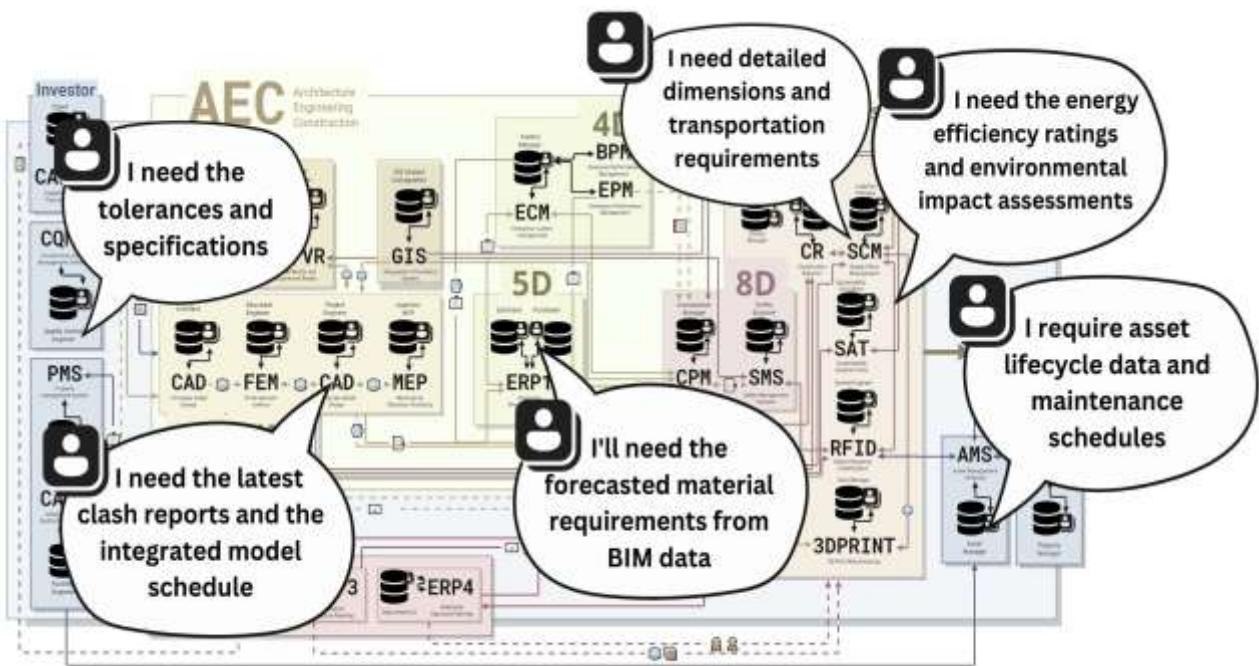
A digitális átjárhatóság a követelményekkel kezdődik

Ahogy a vállalatokon belüli digitális rendszerek száma növekszik, úgy nő az igény az adatok közötti konzisztenciára is. A különböző IT-rendszerekért felelős vezetők gyakran úgy találják, hogy képtelenek lépést tartani a növekvő információmennyiséggel és a különböző formátumokkal. Ilyen körülmények között kénytelenek szakembereket felkérni arra, hogy az adatokat más alkalmazásokban és platformokon való felhasználásra alkalmas formában hozzák létre.

Ez viszont megköveteli, hogy a mérnökök és az adatgeneráló személyzet alkalmazkodjon a követelmények sokaságához, gyakran anélkül, hogy átláthatóságot és világos képet kapnának arról, hogy az adatokat hol és hogyan fogják felhasználni a jövőben. Az információk kezelésére vonatkozó szabványosított megközelítések hiánya eredménytelenséghöz és megnövekedett költségekhez vezet az ellenőrzési szakaszban, amely az adatok összetettsége és nem szabványosított jellege miatt gyakran manuális.

Az adatok szabványosításának kérdése nem csupán kényelmi vagy automatizálási kérdés. Ez közvetlen pénzügyi veszteséget jelent. Egy 2016-os IBM-jelentés szerint a rossz adatminőségből eredő éves veszteség az Egyesült Államokban 3,1 millió dollár [96]. Emellett az MIT és más analitikai tanácsadó cégek tanulmányai azt mutatják, hogy a rossz adatminőség költsége a vállalat bevételének 15-25%-át is elérheti [97].

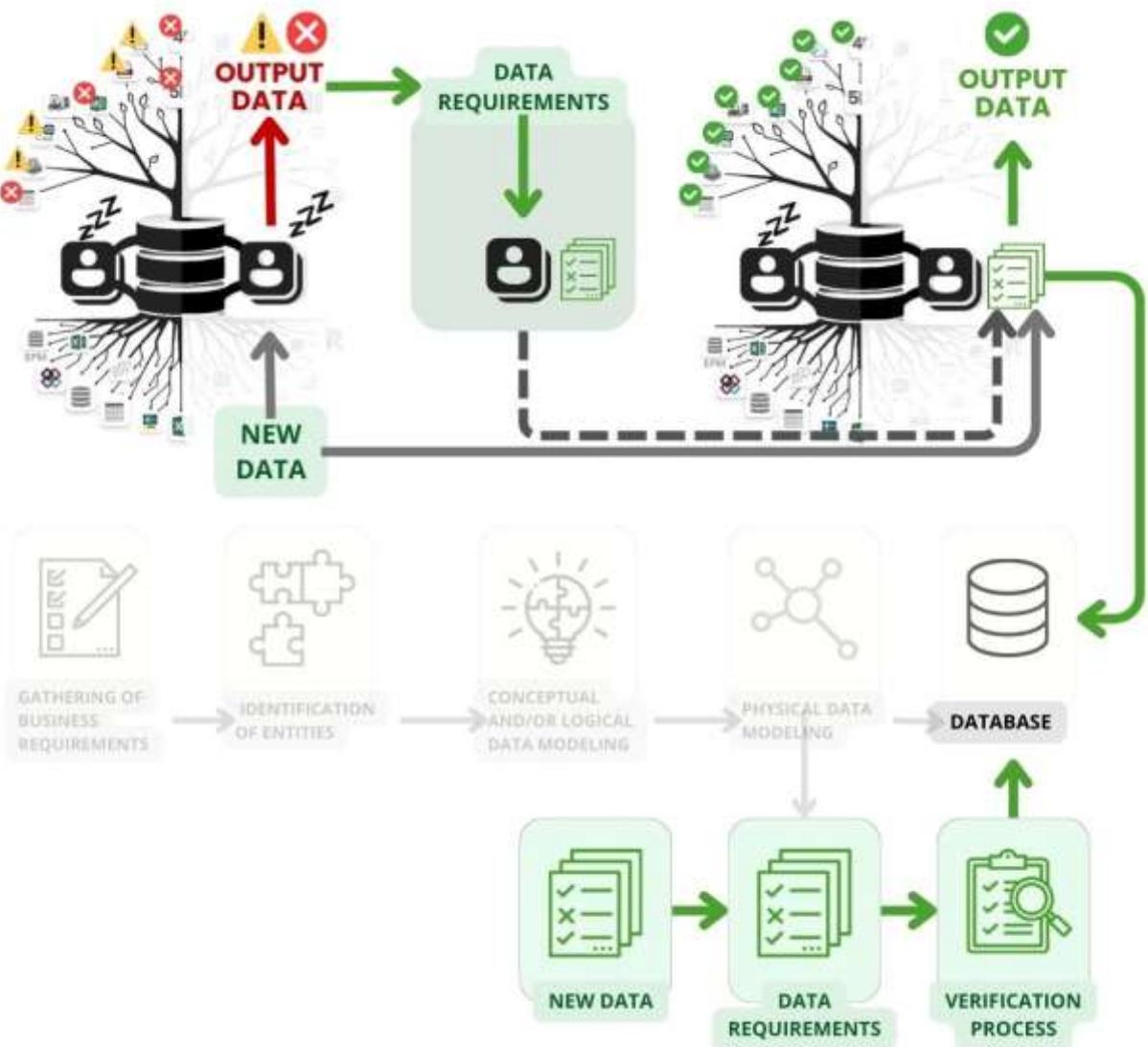
Ilyen körülmények között kritikus fontosságúvá válik, hogy világosan meghatározott adatkövetelményekkel és leírásokkal rendelkezzünk arról, hogy milyen paramétereket, milyen formátumban és milyen részletességgel kell a létrehozott objektumoknak tartalmazniuk. E követelmények formalizálása nélkül lehetetlen garantálni az adatminőséget és a rendszerek és a projektfázisok közötti kompatibilitást (4.2-4. ábra).



4.2-4. ábra Az üzleti tevékenység különböző szerepek kölcsönhatásán alapul, amelyek mindenkorában bizonyos paramétereket és értékeket igényel, amelyek kritikusak az üzleti célok teljesítéséhez.

A megfelelő adatkövetelmények megfogalmazásához, meg kell értenie az üzleti folyamatokat az adatok szintjén. Az építési projektek típusa, nagyságrendje és a résztvevők száma változó, és minden rendszer - legyen szó modellezésről (CAD (BIM)), ütemezésről (ERP 4D), költségszámításról (ERP 5D) vagy logisztikáról (SCM) - saját egyedi paramétereket igényel a bemenetekhez (input entitás-elemek).

Ezektől az igényektől függően az üzleti vezetőknek vagy új adatstruktúrákat kell tervezniük a követelmények teljesítéséhez, vagy a meglévő táblákat és adatbázisokat kell átalakítaniuk. A létrehozott adatok minősége közvetlenül attól függ, hogy a követelményeket mennyire pontosan és helyesen fogalmazták meg (4.2-5. ábra).



4.2-5. ábra Az adatminőség az egyes adatfelhasználási esetekhez létrehozott követelmények minőségétől függ.

Mivel minden rendszernek megvannak a maga sajátos adatigényei, az általános követelmények megfogalmazásának első lépése az üzleti folyamatokban részt vevő összes elem kategorizálása. Ez azt jelenti, hogy az objektumokat osztályokra és osztálycsoportokra kell osztani, amelyek megfelelnek a konkrét rendszereknek vagy alkalmazási feladatoknak. minden ilyen csoportra külön követelményeket kell kidolgozni az adatszerkezetre, az attribútumokra és a minőségre vonatkozóan.

A gyakorlatban azonban e megközelítés megvalósítása komoly kihívással szembesül: az adatok csoportosítására szolgáló közös nyelv hiánya. Az eltérő osztályozások, a duplikált azonosítók és az inkompatibilis formátumok azt eredményezik, hogy minden vállalat, minden softver, sőt minden projekt saját, elszigetelt adatmodelleket és osztályokat hoz létre. Az eredmény egy digitális "Bábel tornya", ahol az információk rendszerek közötti átvitele többszörös, gyakran kézzel végzett konvertálást igényel a megfelelő adatmodellekbe és osztályokba. Ezt az akadályt csak az univerzális osztályozók és szabványosított

követelményrendszerek felé való elmozdulással lehet leküzdeni.

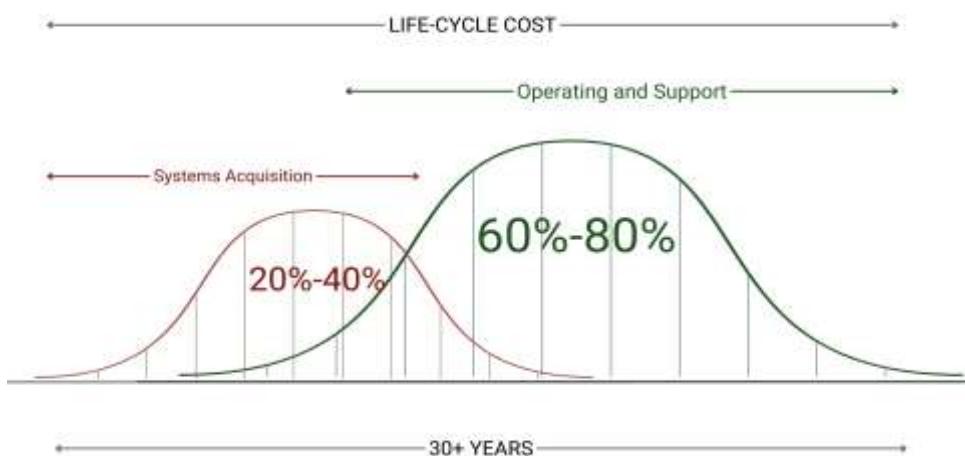
Az építés közös nyelve: az osztályozók szerepe a digitális átalakulásban

Az ellenőrzési és feldolgozási folyamatok digitalizálásával és automatizálásával összefüggésben az osztályozási rendszerek különleges szerepet játszanak - egyfajta "digitális szótárak", amelyek biztosítják az objektumok egységes leírását és paraméterezését. Az osztályozók alkotják azt a "közös nyelvet", amely lehetővé teszi az adatok jelentés szerinti csoportosítását és integrálását a különböző rendszerek, irányítási szintek és a projekt életciklusának különböző fázisai között.

Az osztályozók legkézzelfoghatóbb hatása az épületek életciklusának gazdaságosságában jelentkezik, ahol a legfontosabb szempont a hosszú távú üzemeltetési költségek optimalizálása. Tanulmányok azt mutatják, hogy az üzemeltetési költségek az épület teljes tulajdonlási költségének akár 80%-át is kiteszik, ami háromszor magasabb, mint a kezdeti építési költségek (4.2-6. ábra) [98]. Ez azt jelenti, hogy a jövőbeli költségekre vonatkozó döntés nagyrészt a tervezési szakaszban alakul ki

Ezért az üzemeltetési mérnökötől származó követelmények (CAFM, AMS, PMS, RPM) a tervezési fázisban az adatkövetelmények előállításának kiindulópontjává kell válniuk (1.2-4. ábra). Ezeket a rendszereket nem a projekt végső szakaszának kell tekinteni, hanem a projekt teljes digitális ökoszisztemájának szerves részének, a koncepciótól a lebontásig

A modern osztályozó nem csupán egy csoportosítási kódrendszer. Az építészek, mérnökök, becslők, logisztikusok, karbantartók és informatikai rendszerek közötti kölcsönös megértés mechanizmusa. Ahogyan egy autó robotpilótájának egyértelműen és pontosan fel kell ismernie az úttest objektumait, a digitális építési rendszereknek és felhasználóknak az elemosztályon keresztül ugyanazt a projektelemet a különböző rendszerek számára egyértelműen kell értelmezniük.

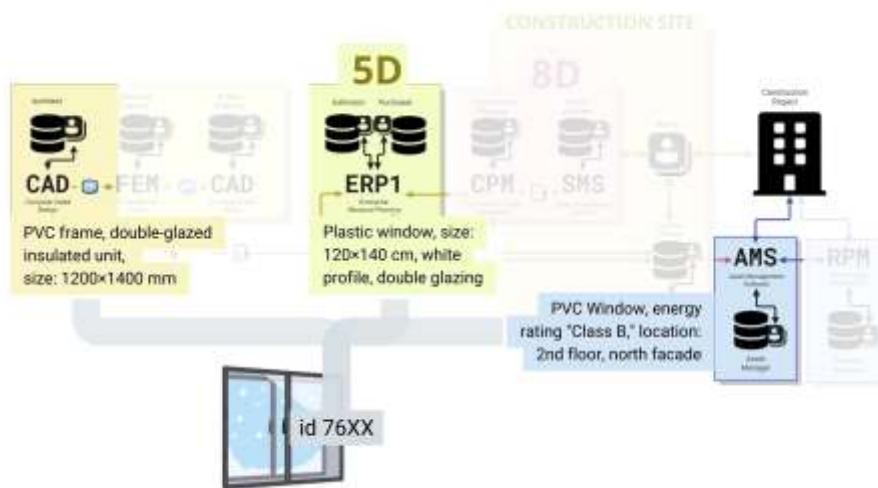


ábra 4.2-6 Az üzemeltetési és karbantartási költségek háromszorosan meghaladják az építési költségeket, és az épület teljes életciklusköltségének 60-80%-át teszik ki ([99] szerint).

Az osztályozók fejlesztésének szintje közvetlenül korrelál a vállalat digitalizációjának mélységével és digitális érettségével. Az alacsony szintű digitális érettséggel rendelkező szervezetek széttöredezett

adatokkal, inkompatibilis információs rendszerekkel és ennek következtében inkompatibilis és nem hatékony osztályozókkal szembesülnek. Az ilyen vállalatoknál ugyanaz az elem gyakran különböző csoportosítási azonosítókkal rendelkezhet a különböző rendszerekben, ami kritikusan akadályozza a végső integrációt és lehetetlenné teszi a folyamatok automatizálását.

Például egy projekt ugyanazon ablakát a CAD modellben, a becslési és karbantartási rendszerben (4.2-7. ábra) különbözőképpen lehet címkézni, mivel a folyamat különböző résztvevői az elemeket többdimenziósan érzékelik. A becslő számára az ablak kategóriában az elem, a mennyiség és a költség, a karbantartó szolgálat számára a rendelkezésre állás és a karbantarthatóság, az építész számára az esztétikai és funkcionális jellemzők fontosak. Ennek eredményeképpen ugyanaz az elem különböző paramétereket igényelhet.



4.2-7 A rendszerek közötti következetlen osztályozás esetén egy elem a másik rendszerbe való átmenet minden egyes lépésénél elveszít bizonyos attribútuminformációkat.

Mivel az épületelemek osztályozását nehéz egyértelműen meghatározni, a különböző szakterületek szakemberei gyakran nem összeegyeztethető osztályokat rendelnek ugyanahhoz az elemhez. Ez az objektumról alkotott egységes kép elvészítéséhez vezet, ami utólagos kézi beavatkozást igényel a különböző osztályozási rendszerek harmonizálása és a különböző szakemberek által meghatározott típusok és osztályok közötti összhang megteremtése érdekében.

E következetlenség következtében a beszerzési osztály (ERP) által egy építési térel gyártótól történő beszerzésekor kapott operatív dokumentáció nem kapcsolható össze helyesen az adott térel építési helyszínén történő besorolásával (PMIS, SCM). Ennek következtében a kritikus információk valószínűleg nem kerülnek integrálásra az infrastruktúra- és eszközkezelési rendszerekbe (CAFM, AMS), ami komoly problémákat okoz az üzembe helyezés során, valamint a későbbi karbantartás (AMS, RPM) vagy az elem cseréje során.

A magas digitális érettségű vállalatoknál az osztályozók az összes információáramlást egyesítő idegrendszer szerepét töltik be. Ugyanaz az elem egyedi azonosítót kap, így torzítás és veszteség nélkül továbbítható a CAD, ERP, AMS és CAFM-rendszer és osztályozóik között.

A hatékony osztályozók létrehozásához meg kell értenie, hogyan használják az adatokat. Ugyanaz a mérnök

különböző projektekben másképp nevezhet el és osztályozhat egy elemet. Csak az évek során gyűjtött használati statisztikák segítségével lehet stabil osztályozási rendszert kialakítani. Ebben segít a gépi tanulás: az algoritmusok projektek ezreit elemzik (9.1-10. ábra), és gépi tanulással azonosítják a valószínű osztályokat és paramétereket (10.1-6. ábra). Az automatikus osztályozás különösen értékes olyan környezetben, ahol a kézi osztályozás az adatmennyiség miatt nem lehetséges. Az automatikus osztályozó rendszerek képesek lesznek az alapvető kategóriák megkülönböztetésére a minimálisan feltöltött elemparaméterek alapján (további részletek a könyv kilencedik és tizedik részében).

A kifejlesztett osztályozórendszerek a további digitalizálás katalizátorává válnak, megteremtve az alapot:

- A projektköltségek és határidők automatizált becslése.
- A potenciális kockázatok és konfliktusok előrejelző elemzése
- A beszerzési folyamatok és a logisztikai láncok optimalizálása
- Épületek és szerkezetek digitális ikeinek létrehozása
- Integráció az intelligens városi és a tárgyak internetét használó rendszerekkel

Az átalakulás ideje korlátozott - a gépi tanulás és a számítógépes látás technológiáinak fejlődésével az automatikus osztályozás évtizedek óta megoldhatatlan problémája az elkövetkező években megoldódik, és azok az építőipari és tervező cégek, amelyek nem alkalmazkodnak időben, azt kockáztatják, hogy megismétlődik a digitális platformok által kiszorított taxiflották sorsa.

A költség- és ütemterv-számítások automatizálását, valamint a nagy adatmennyiséget és a gépi tanulást a könyv ötödik és kilencedik részében tárgyaljuk részletesebben. A taxiflották sorsának megismétlődésének kockázatát és az építőipar Uberizációját a könyv tizedik része tárgyalja részletesen.

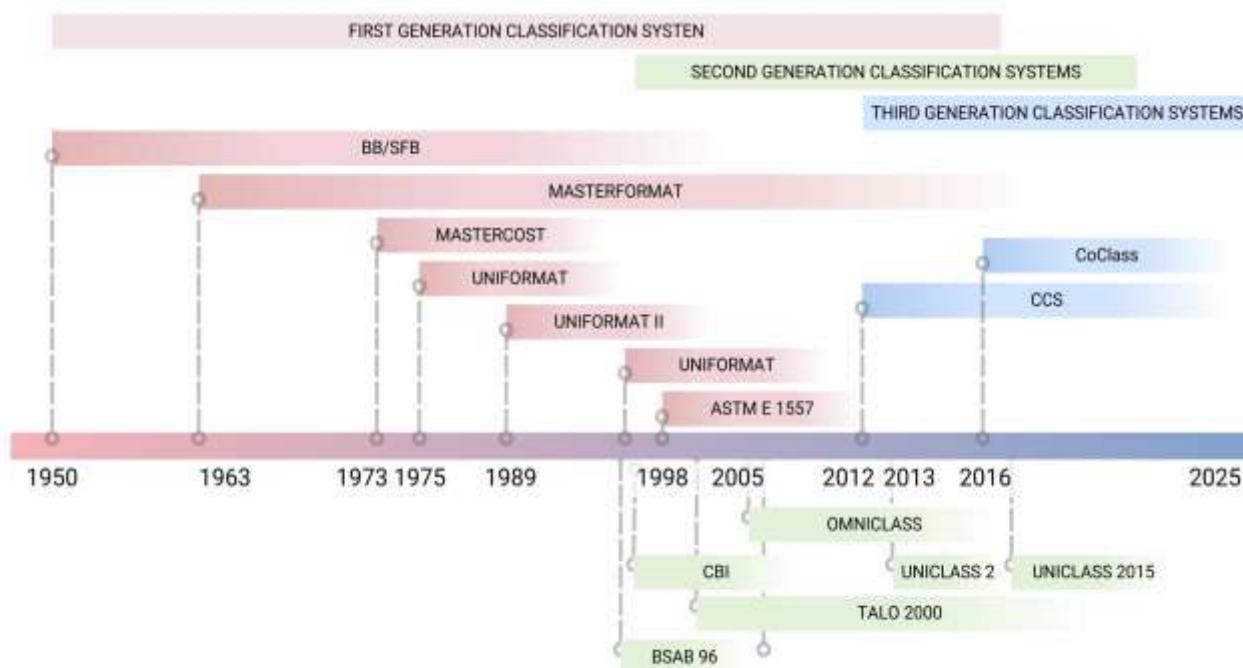
Az osztályozóknak az építőipar digitális átalakulásában betöltött kulcsszerepének megértéséhez szükséges, hogy áttekintsük fejlődésük történetét. A történelmi kontextus az, amely lehetővé teszi számunkra, hogy felismerjük, hogyan fejlődtek az osztályozással kapcsolatos megközelítések, és milyen trendek határozzák meg jelenlegi állapotukat.

Masterformat, OmniClass, Uniclass és CoClass: az osztályozási rendszerek fejlődése

Történelmileg az építőipari elemek és munkák osztályozói három generációban fejlődtek, amelyek mindegyike a rendelkezésre álló technológia szintjét és az iparág aktuális igényeit tükrözte egy adott időszakban (4.2-8. ábra):

- **Első generáció** (1950-es évek elejétől az 1980-as évek végéig) - papíralapú könyvtárak, helyben használt hierarchikus osztályozók (pl. Masterformat, SfB).
- **A második generáció** (1990-es évek vége-2010-es évek közepe) az Excel és Access programokban megvalósított táblázatkezelő táblázatok és strukturált adatbázisok (ASTM E 1557, OmniClass, Uniclass 1997).
- **Harmadik generáció** (2010-es évektől napjainkig) - digitális szolgáltatások és API-k -interfészek,

integráció a CAD (BIM), automatizálás (Uniclass 2015, CoClass).



Ábra 4.2-8 Az építőipari osztályozók három generációja.

Az elmúlt évtizedekben csökkent az osztályozók hierarchikus komplexitása (4.2-9. ábra): míg a korai rendszerek, mint például az OmniClass, akár 7 fészkelődési szintet is használtak 6887 osztály leírására, a modern megoldások, mint például a CoClass, 3 szintre korlátozódnak 750 osztállyal. Ez megkönnyíti az adatokkal való munkát a szükséges granularitás fenntartása mellett. Az Egyesült Királyságban gyakran szabvánnyként használt Uniclass 2015 mindenben 7210 osztályt egyesít, ami kényelmessé teszi a CAD projekteket és a közbeszerzések számára.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Ábra 4.2-9 Az osztályozók minden egyes új generációjával a kategorizálás bonyolultsága drámaian csökken.

A különböző országok építési becslési rendszereiben még egy tipikus elem, például egy beton alapfal is nagyon különböző módon írható le a különböző osztályozások miatt (4.2-10. ábra). Ezek a különbözők a nemzeti építési gyakorlatokat, az alkalmazott mérési rendszereket, az anyagok osztályozásának

megközelítéseit, valamint az egyes országokban hatályos szabályozási és műszaki követelményeket tükrözik.



Ábra 4.2-10 Ugyanaz az elem a különböző országok projektjeiben különböző leírások és besorolások révén kerül felhasználásra.

Az azonos elemek osztályozásának sokfélesége megnehezíti a nemzetközi együttműködést, és időigényessé, olykor szinte lehetetlennek teszi a költségek és a munkaterület összehasonlítását a nemzetközi projekteken belül. Jelenleg nincs egyetlen egyetemes, globális szintű osztályozási rendszer - minden ország vagy régió saját rendszereket dolgoz ki a helyi normák, nyelv és üzleti kultúra alapján:

- **CCS** (Dánia): Költségosztályozási rendszer - a költségek osztályozására szolgáló rendszer a létesítmény teljes életciklusa során (tervezés, építés, üzemeltetés). A hangsúly az üzemeltetési és karbantartási logikán van, de a költségvetési és erőforrás-gazdálkodásra is kiterjed.
- **NS 3451** (Norvégia): a létesítményeket funkció, tervezési elemek és életciklus-fázisok szerint kategorizálja. Projektmenedzsmenthez, költségbecsléshez és hosszú távú tervezéshez használatos.
- **MasterFormat** (USA): az építési előírások szakaszokra (pl. beton, elektromos, kivitelezés) történő strukturálására szolgáló rendszer. A funkcionális elemek helyett inkább a szakterületekre és a munkatípusokra összpontosít (ellentétben az UniFormattal).
- **Uniclass 2** (UK): az egyik legrészletesebb osztályozó, amelyet a közbeszerzési és BIM projektekben használnak. Egységesíti a tárgyakra, munkákra, anyagokra és terekre vonatkozó adatokat egyetlen rendszerben.
- **OmniClass**: nemzetközi szabvány (amelyet az amerikai CSI fejlesztett ki) a komponenskönyvtárakból származó objektuminformációk elektronikus specifikációigig történő kezelésére. Alkalmas hosszú távú adattárolásra, kompatibilis a CAD (BIM) és más digitális eszközökkel.
- **COBie**: Construction-Operation Building information exchange (építési-üzemeltetési épületinformációcsere) a tervezési, építési és üzemeltetési fázisok közötti adatcsere nemzetközi szabványa. A BS 1192-4:2014 szabványban szerepel a "BIM - használatra kész modell" koncepció részeként. Az információátadásra összpontosít (pl. berendezésekre vonatkozó előírások, garanciák, kivitelezői kapcsolatok).

Az építőipar globalizációja valószínűleg az építőelemek osztályozási rendszereinek fokozatos egységesítéséhez vezet, ami jelentősen csökkenteni fogja a helyi nemzeti szabványoktól való függőséget. Ez a folyamat az internetes kommunikáció fejlődéséhez hasonlóan alakulhat, ahol az egyetemes adatátviteli protokollok végül kiszorították az eltérő helyi formátumokat, biztosítva a rendszerek globális

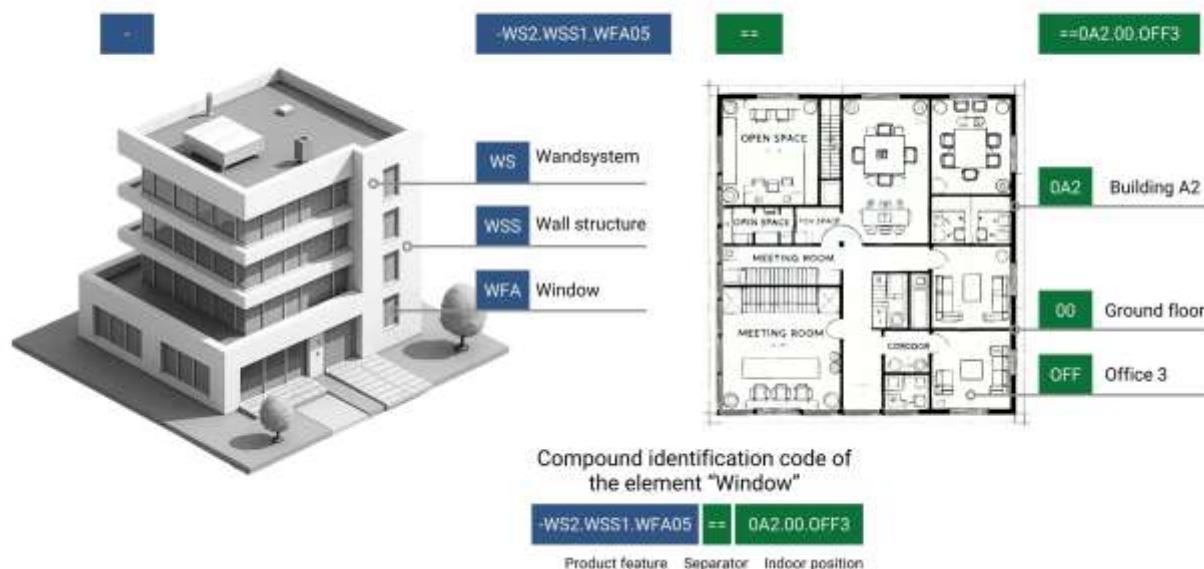
interoperabilitását.

Egy alternatív fejlesztési útvonal lehet a gépi tanulási technológiákon alapuló automatikus osztályozási rendszerekre való közvetlen áttérés. Ezek a technológiák, amelyeket ma elsősorban az autonóm közlekedés területén fejlesztenek, jelentős potenciállal rendelkeznek a nagy CAD tervezési adathalmazokra való alkalmazásban (10.1-6. ábra).

A helyzet ma már nem csak az osztályozók nemzeti klaszterezésére korlátozódik. A nemzeti szinten figyelembe nem vett számos sajátosság miatt minden egyes vállalatnak egységesítenie és szabványosítania kell az elemek és erőforrások kategóriáit, amelyekkel dolgozik.

Ez a folyamat általában kicsiben kezdődik - az objektumok helyi táblázataival vagy belső címkézési rendszerekkel. A stratégiai cél azonban az, hogy az összes elem leírására egy olyan közös nyelvre térjünk át, amely nemcsak a vállalaton belül, hanem azon kívül is érhető - ideális esetben nemzetközi vagy iparági osztályozókkal harmonizálva (4.2-8. ábra). Ez a megközelítés megkönnyíti a külső partnerekkel és a digitális rendszerekkel való integrációt, és elősegíti az egységes végponttól végpontig tartó folyamatok kialakítását az objektumok életciklusán belül.

Az automatizálás és a skálázható informatikai rendszerek felé történő elmozdulás előtt vagy nemzeti szintű osztályozókat kell használni, vagy saját, logikus és egyértelmű elemazonosítási struktúrát kell létrehozni. minden objektumot - legyen az egy ablak (4.2-11. ábra), egy ajtó vagy egy gépészeti rendszer - úgy kell leírni, hogy az bármely vállalat digitális rendszerében félreérthetetlenül felismerhető legyen. Ez kritikus fontosságú a síkbeli rajzokról a digitális modellekre való áttérés során, amely minden a tervezési fázisra, minden az épület üzemeltetésére kiterjed.



ábra 4.2-11 Példa az összetett ablakok épületelem-azonosítójára a osztályozás és az épületben elfoglalt hely alapján.

A belső osztályozók egyik példája lehet egy összetett azonosító kód kifejlesztése (4.2-11. ábra). Egy ilyen kód több szintű információt egyesít: az elem funkcionális rendeltetését (pl. "ablak a falban"), típusát és pontos térbeli vonatkozásait - A2 épület, 0. emelet, 3. szoba. Az ilyen többszintű struktúra lehetővé teszi a digitális modellekben és dokumentációban való egységes navigációs rendszer létrehozását, különösen az adatok ellenőrzésének és átalakításának szakaszában, ahol az elemek egyértelmű csoportosítására van szükség. Az elemek egyértelmű felismerése biztosítja a részlegek közötti összhangot, és csökkenti a duplikációk, hibák és információvesztés kockázatát.

Egy jól felépített osztályozó nem csupán egy technikai dokumentum, hanem a vállalat digitális ökoszisztémájának alapja:

- biztosítja az adatok kompatibilitását a rendszerek között;
- csökkenti az információk keresésének és feldolgozásának költségeit;
- növeli az átláthatóságot és a kezelhetőséget;
- megteremti a méretezés és az automatizálás alapját.

Az objektumok szabványosított leírása a nemzeti osztályozók vagy saját összetett azonosító kódok alkalmazásával a következő adatok, a megbízható információcsere és az intelligens szolgáltatások későbbi megvalósításának alapjává válik - az automatizált beszerzéstől a digitális ikrekig.

A több formátumú adatok strukturálási szakaszának befejezése és az elemek felismerésére és csoportosítására használt osztályozó kiválasztása után a következő lépés az adatok megfelelő modellezése. Ez a folyamat magában foglalja a kulcsfontosságú paraméterek azonosítását, a logikus



adatszerkezet felépítését és az elemek közötti kapcsolatok leírását.

FEJEZET 4.3.

ADATMODELLEZÉS ÉS KIVÁLÓSÁGI KÖZPONT

Adatmodellezés: fogalmi, logikai és fizikai modell

A (korábban általunk strukturált és kategorizált) adatok hatékony kezelése lehetetlen egy jól átgondolt tárolási és feldolgozási struktúra nélkül. Az információhoz való hozzáférés és az információk konzisztenciájának biztosítása érdekében a tárolási és feldolgozási szakaszokban a vállalatok adatmodellezést alkalmaznak, egy olyan módszertant, amely lehetővé teszi számukra a táblázatok, adatbázisok és a közöttük lévő kapcsolatok üzleti követelményeknek megfelelő kialakítását.

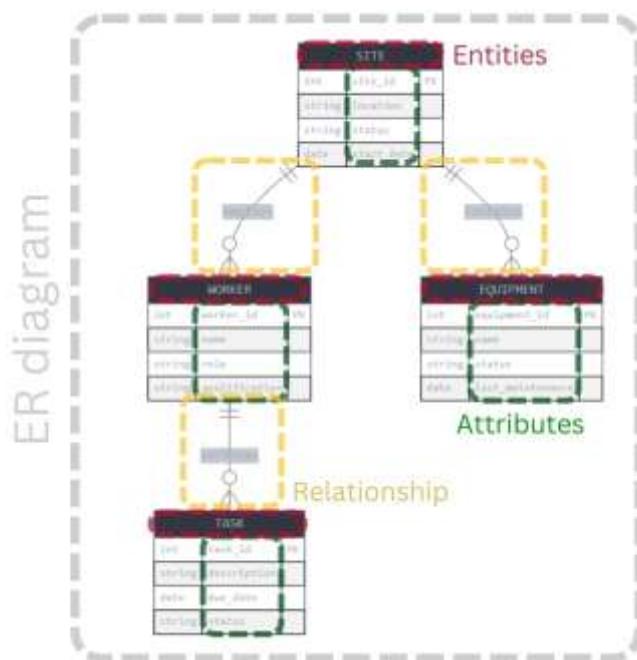
Az adatmodellezés az alapja minden digitális ökoszisztemának. A rendszerek, a követelmények és az adatmodellezés leírása nélkül az adatokat létrehozó mérnökök és szakemberek nem tudják vagy nem értik, hogy az általuk létrehozott adatokat hol fogják felhasználni.

Mint egy épület építése, ahol terv nélkül nem lehet elkezdeni a téglák lerakását, az adattárház-rendszer létrehozásához világos elképzelés szükséges arról, hogy milyen adatokat fognak használni, hogyan fogják összekapcsolni, és ki fog vele dolgozni. A folyamatok és követelmények leírása nélkül az adatokat létrehozó mérnökök és szakemberek szem elől tévesztik, hogy az adatokat hol és hogyan fogják használni a jövőben.

Az adatmodell hídként szolgál az üzlet és az IT között. Lehetővé teszi a követelmények formalizálását, az információk strukturálását és az érdekkelt felek közötti kommunikáció megkönnyítését. Ebben az értelemben az adatmodellezés az építész munkájához hasonlít, aki a megrendelő tervei alapján kidolgozza az épület tervét, majd továbbadja azt az építőknek - adatbázis-adminisztrátoroknak és fejlesztőknek - a megalósításhoz (adatbázis létrehozása).

Így minden építőipari vállalatnak az elemek és erőforrások strukturálásán és kategorizálásán (4.2-11. ábra) túl el kell sajátítania az adatbázisok (táblázatok) "építésének" művészetét, és meg kell tanulnia, hogyan hozzon létre kapcsolatokat közöttük, mintegy összekötve a téglákat a vállalat adataiból származó tudás megbízható és erős falává. Az adatmodellezés kulcsfogalmai (4.3-1. ábra) a következők:

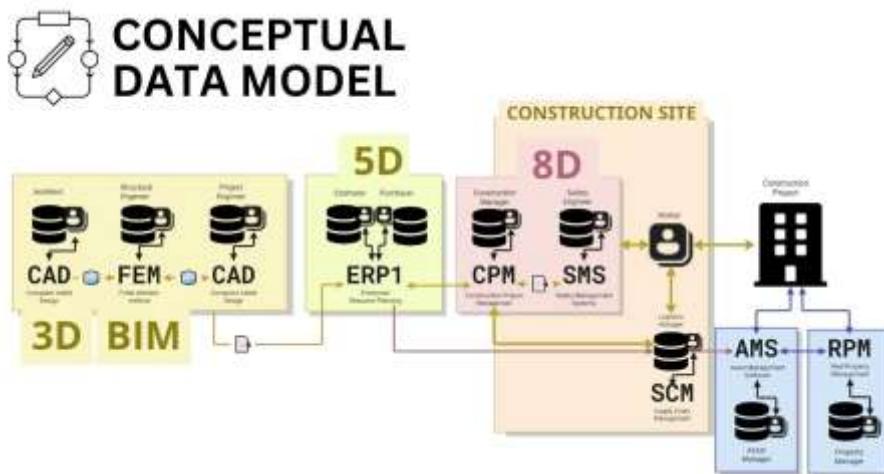
- **Az entitások** olyan objektumok, amelyekről adatokat kell gyűjteni. A korai tervezési fázisban egy entitás lehet egyetlen elem (pl. "ajtó"), a becslési modellben pedig kategorizált elemek csoportja (pl. "belső ajtók").
- **Az attribútumok** az entitások olyan jellemzői, amelyek fontos részleteket írnak le: méretek, tulajdonságok, összeszerelési költségek, logisztika és egyéb paraméterek.
- **Kapcsolatok (linkek)** - megmutatják, hogy az entitások hogyan hatnak egymásra. A kapcsolatok a következő típusok valamelyike lehetnek: "egy az egyhez", "sok az egyhez", "sok a sokhoz".
- Az ER-diagramok (Entity-Relationship diagramok) olyan vizuális diagramok, amelyek a entitásokat, attribútumokat és a köztük lévő kapcsolatokat mutatják. Az ER-diagramok lehetnek fogalmi, logikai és fizikai diagramok - mindegyik különböző részletességi szintet tükröz.



4.3-1. ábra Egy fogalmi adatbázis szerkezetének ER-diagramja entitásokkal, attribútumokkal és kapcsolatokkal.

Az adatok tervezésének és a köztük lévő kapcsolatok meghatározásának folyamatát hagyományosan három fő modellre osztják. Ezek mindegyike bizonyos funkciókat lát el, és az adatszerkezet ábrázolásának részletességeiben és absztrakciós fokában különböznek:

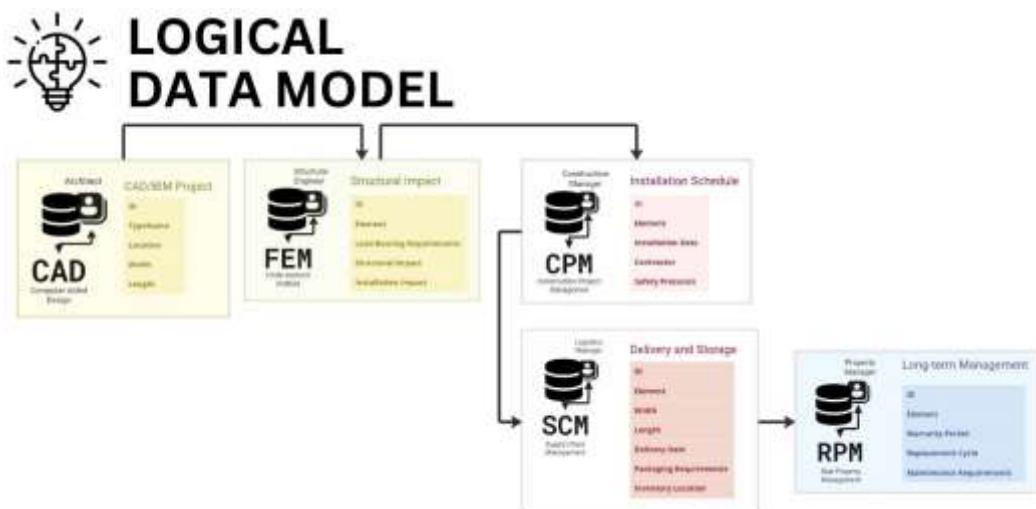
- **Fogalmi adatmodell:** ez a modell a fő entitásokat és kapcsolataikat írja le anélkül, hogy az attribútumok részleteire kiterne. Általában a tervezés kezdeti szakaszában használják. Ebben a szakaszban adatbázisokból és rendszerekből vázlatokat készíthetünk a különböző részlegek és szakemberek közötti kapcsolatok bemutatására.



Ábra 4.3-2 A koncepcionális diagram a rendszer tartalmát írja le: a kapcsolatok magas szintű ábrázolása, technikai részletek nélkül.

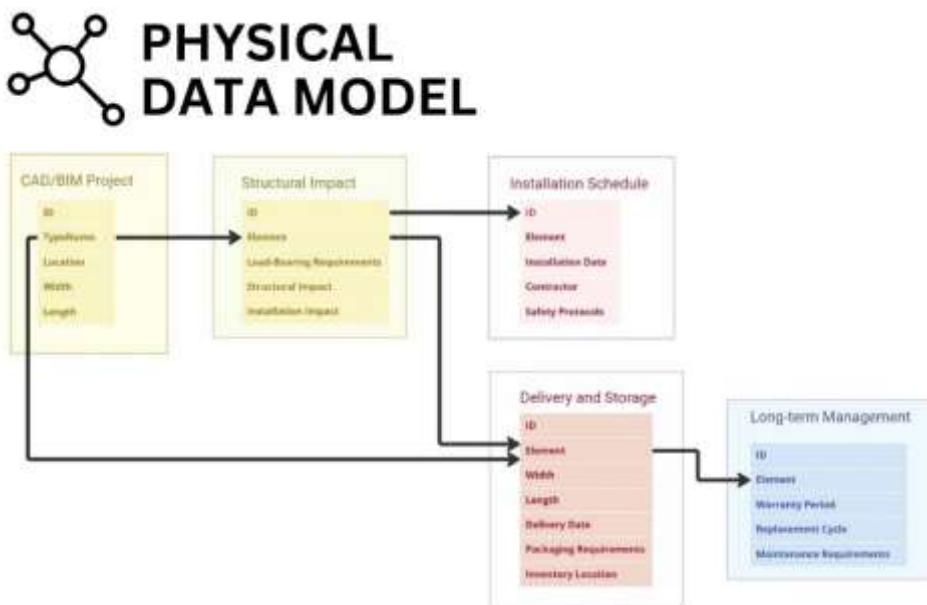
- **Logikai adatmodell:** A fogalmi modell alapján a logikai adatmodell tartalmazza az entitások,

attribútumok, kulcsok és kapcsolatok részletes leírását, az üzleti információk és szabályok leképezését.



A 4.3-3. ábra Logikai adatmodell részletesen, de a rendszer implementálása nélkül írja le az adattípusokat, kapcsolatokat és kulcsokat.

- Fizikai adatmodell:** Ez a modell leírja az adatbázis megvalósításához szükséges struktúrákat, beleértve a táblákat, oszlopotokat és kapcsolatokat. Az adatbázisok fizikai telepítésének optimalizálása érdekében az adatbázisok teljesítményére, az indexelési stratégiákra és a fizikai tárolásra összpontosít.

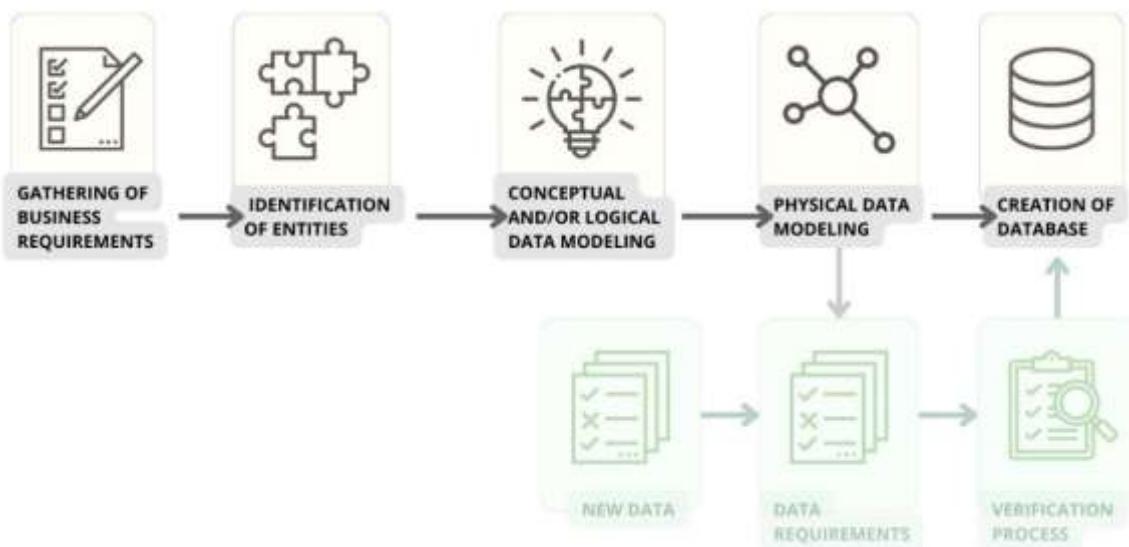


4.3-4. ábra A fizikai adatmodell meghatározza, hogy a rendszer hogyan lesz megvalósítva, beleérte a táblákat és az adatbázis konkrét részleteit.

Az adatbázisok tervezésekor és a táblázatos kapcsolatok kialakításakor az absztrakciós szintek megértése kulcsszerepet játszik a hatékony rendszerarchitektúra kialakításában.

A hatékony adatmodellezési módszertan lehetővé teszi az üzleti célok és a technikai megvalósítás összekapcsolását, így a teljes folyamatlánc átláthatóbbá és kezelhetőbbé válik. Az adatmodellezés nem egy egyszeri feladat, hanem egy folyamat, amely egymást követő lépésekkel áll (4.3-5. ábra):

- **Az üzleti követelmények összegyűjtése:** a legfontosabb feladatak, célkitűzések és információáramlások meghatározása. Ez a szakértőkkel és a felhasználókkal való aktív interakció szakasza.
- **Entitás azonosítása:** kiemelésre kerülnek a fő objektumok, kategóriák és adattípusok, amelyeket fontos figyelembe venni a jövőbeli rendszerben.
- **Fogalmi és logikai modell kidolgozása:** először a kulcsfontosságú entitások és kapcsolataik, majd az attribútumok, szabályok és a részletes struktúra rögzítése.
- **Fizikai modellezés:** a modell technikai megvalósításának megtervezése: táblák, mezők, kapcsolatok, korlátozások, indexek.
- **Adatbázis létrehozása:** az utolsó lépés a fizikai modell megvalósítása a kiválasztott DBMS-ben, a tesztelés elvégzése és az üzembe helyezés előkészítése.



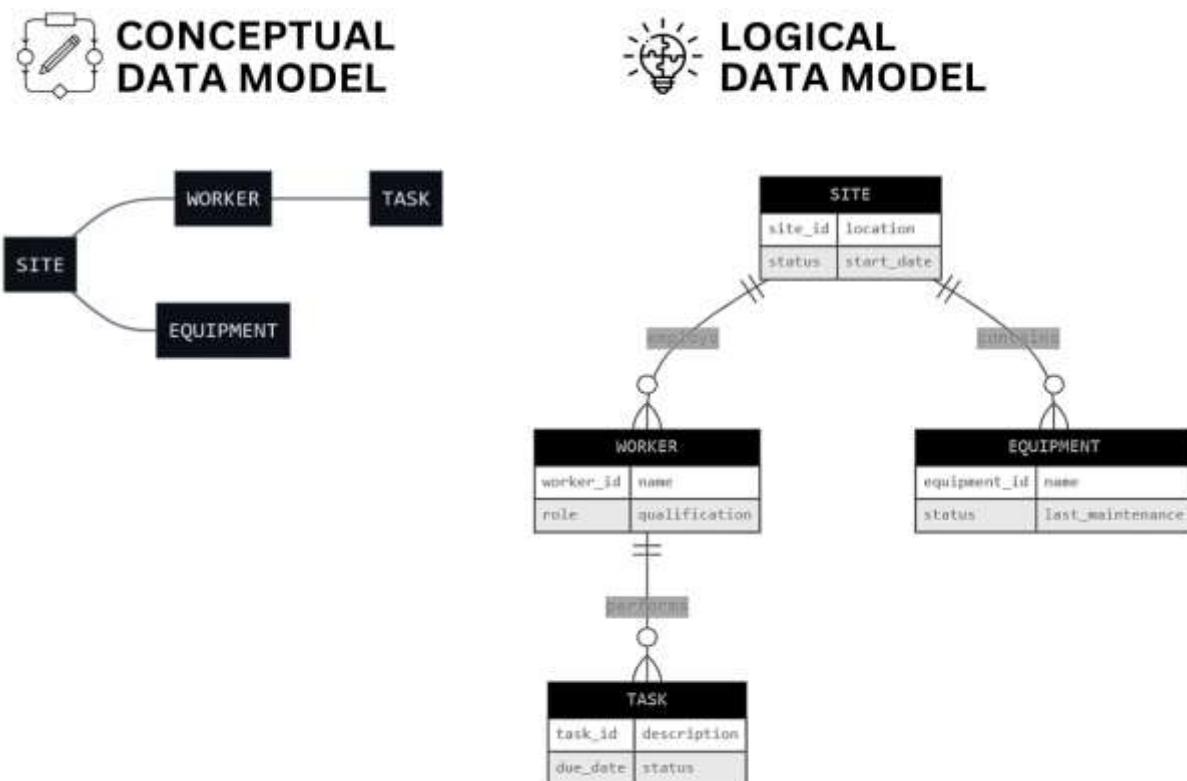
Ábra: 4.3-5 Az üzleti folyamatokhoz szükséges adatbázisok és adatkezelő rendszerek létrehozása a követelmények létrehozásával és az adatok modellezésével kezdődik.

A megfelelően megtervezett adatmodellezési folyamatok átlátható információáramlást tesznek lehetővé, ami különösen fontos az olyan összetett projektek esetében, mint az építési projektek vagy a telephelyek irányítása. Nézzük meg, hogy a koncepcionális modellről a logikai modellre, majd a fizikai modellre való áttérés hogyan segíthet a folyamatok racionalizálásában.

Gyakorlati adatmodellezés az építőiparban

Vegyük egy építkezés irányítási feladatot az adatmodellezés példájaként, és alakítsuk át a művezető követelményeit strukturált logikai modellé. Az építkezés irányításának alapvető igényei alapján meghatározzuk a következő kulcs entitásokat: építkezés (SITE), munkások (WORKER), felszerelés (EQUIPMENT), feladatok (TASK) és felszereléshasználat (EQUIPMENT_USAGE). minden entitás tartalmaz egy sor attribútumot, amelyek fontos jellemzőket tükröznek. Például a TASK esetében ez lehet a feladat leírása, az esedékesség dátuma, státusza, prioritása; a WORKER esetében pedig a neve, a telephelyen betöltött szerepe, jelenlegi foglalkoztatása stb.

A logikai modell kapcsolatot teremt ezen entitások között, megmutatva, hogy a valós munkafolyamatokban hogyan hatnak egymásra (4.3-6. ábra). Például a telephely és a munkavállalók közötti kapcsolat azt jelzi, hogy egy telephelyen sok munkavállaló dolgozhat, míg a munkavállalók és a feladatok közötti kapcsolat azt tükrözi, hogy egy munkavállaló több feladatot is elvégezhet.

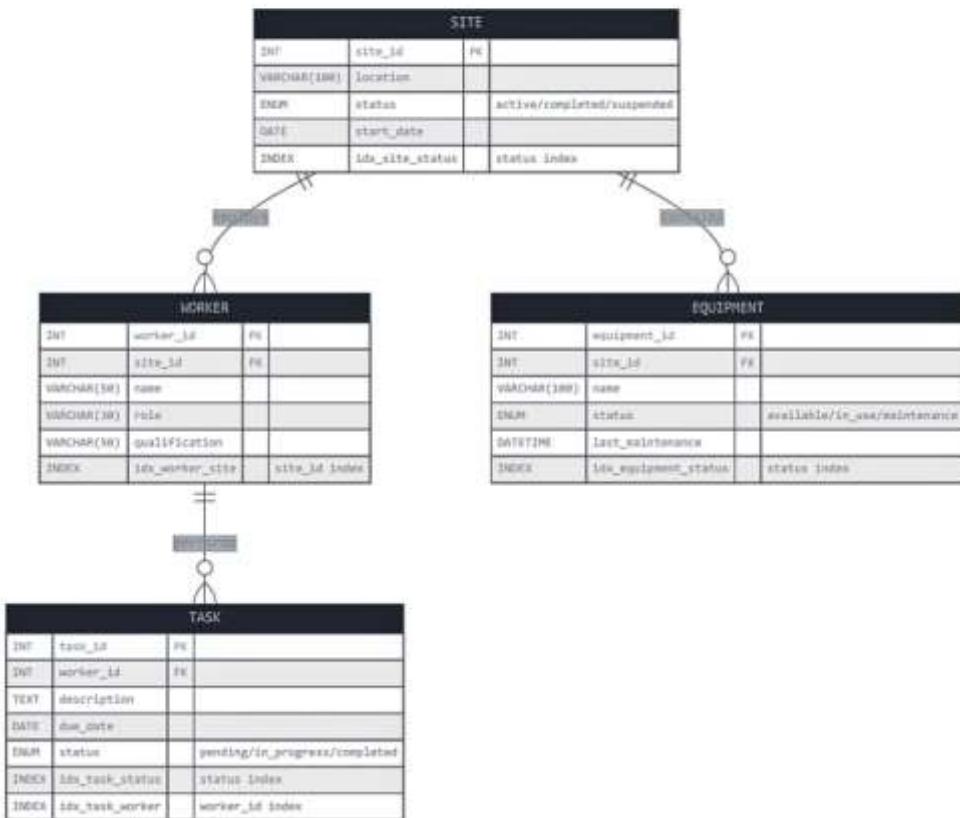


4.3-6. ábra A művezető követelményei alapján generált fogalmi és logikai adatmodell az építkezési folyamatok leírására.

A fizikai modellre való áttéréskor a technikai megvalósítás részletei kerülnek hozzá: speciális adattípusok (VARCHAR, INT, DATE), elsődleges és idegen kulcsok a táblák közötti kapcsolatokhoz, valamint indexek az adatbázis teljesítményének optimalizálása érdekében (4.3-7. ábra).

Például a státuszok számára specifikus típusokat kell meghatározni a lehetséges értékekkel, és a keresési teljesítmény javítása érdekében indexeket kell hozzáadni a kulcsmezőkhöz, például a státusz és a worker_id mezőkhöz. Így a rendszer logikai leírása konkrét adatbázis-megvalósítási tervválik, amely készen áll a létrehozásra és megvalósításra.

PHYSICAL DATA MODEL



4.3-7 ábra A fizikai adatmodell az építkezéshez szükséges minimális paraméterek segítségével írja le az építkezés egységeit.

A fizikai modell gyakran eltér a logikai modelltől. A modellezésre fordított idő átlagosan a következőképpen oszlik meg: körülbelül 50%-át a koncepcionális modellel (követelmények összegyűjtése, folyamatok megvitatása, entitások azonosítása), 10%-át a logikai modellel (attribútumok és kapcsolatok meghatározása) és 40%-át a fizikai modellel (megvalósítás, tesztelés, DBMS-hez való igazítás) töltik.

Ez az egyensúly annak köszönhető, hogy a fogalmi szakasz megalapozza az adatszerkezetet, míg a logikai modell csak a kapcsolatokat és az attribútumokat határozza meg. A fizikai modell igényli a legtöbb erőforrást, mivel ebben a szakaszban kerül sor az adatok konkrét platformokra és eszközökre történő implementálására.

Adatbázis létrehozása az LLM segítségével

Az adatmodell és az entitások paramétereeken keresztül történő leírása után készen állunk az adatbázisok - tárolók - létrehozására, ahol a strukturálási szakasz után a konkrét folyamatokra vonatkozó információkat fogjuk tárolni.

Próbálunk meg egy egyszerű, de funkcionális adatbázist létrehozni minimális mennyiséggű kóddal az SQLite segítségével a Python programozási nyelv példáján keresztül. A relációs adatbázisokat a "Strukturált relációs adatbázisok és az SQL lekérdezőnyelv" című fejezetben részletesen tárgyaltuk.

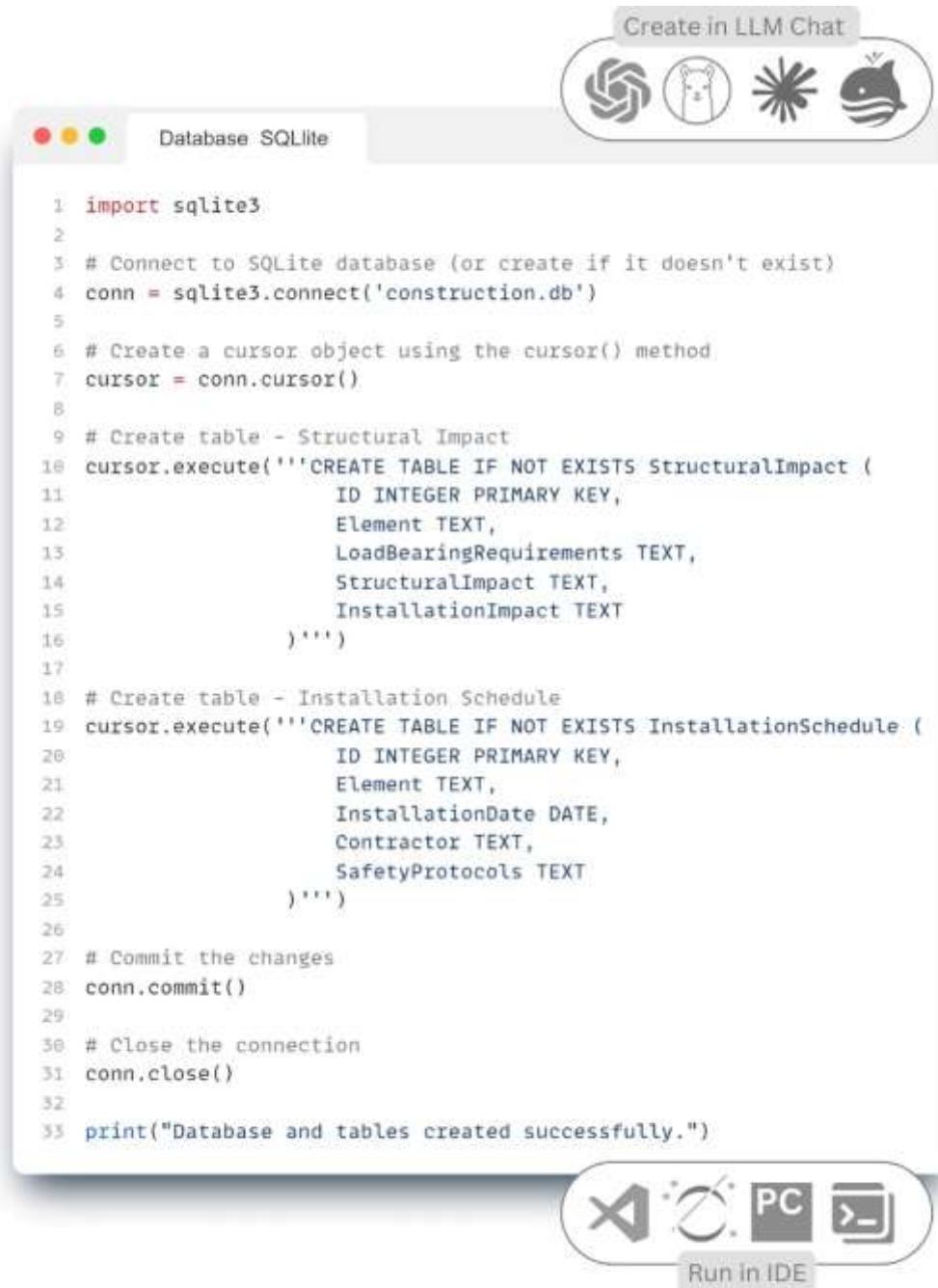
Az SQLite egy könnyű, nyílt forráskódú, beágyazható, keresztplatformos adatbázis, amely támogatja az alapvető SQL funkciókat. Az SQLite-ot széles körben használják a CAD (BIM), ERP, EPM és szinte minden okostelefonban megtalálható.

Kérjük meg az LLM modellt, hogy a fent létrehozott fizikai modell (4.3-4. ábra) alapján hozzon létre számunkra egy kis SQLite adatbázist a FEM (szerkezeti hatás) és a CPM (szerelési ütemterv) rendszerek számára.

- ❷ Küldjön szöveges kérést az LLM chat (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vagy bármely más):

Hozzon létre SQLite adatbázisokat a "Constructive Impact" és a "Installation Schedule" rendszerekhez a csatolmányban található ábra alapján (illeszze be a sémát képként a chatbe (4.3-4. ábra) ↴

■ LLM válasza :



```

1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11                 ID INTEGER PRIMARY KEY,
12                 Element TEXT,
13                 LoadBearingRequirements TEXT,
14                 StructuralImpact TEXT,
15                 InstallationImpact TEXT
16             )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20                 ID INTEGER PRIMARY KEY,
21                 Element TEXT,
22                 InstallationDate DATE,
23                 Contractor TEXT,
24                 SafetyProtocols TEXT
25             )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")

```




Run in IDE

4.3-8 ábra A generált kód a feltöltött képből adatbázisokat és táblákat hoz létre, amelyekbe most már betölthetjük az adatokat.

Ez az adatbázis kód futtatható a Python környezetben az SQLite támogatással, amely általában a standard Python környezet része. Bármely helyi Python gépen futtatható, amely képes szkriptek futtatására és fájlok létrehozására. A futtatás és végrehajtás után ez a kód létrehoz egy fájlt a merevlemezen - egy SQLite adatbázist nevű construction.db (hasonlóan egy Excel fájlhoz munkalapokkal), amely a "Szerkezeti hatás" és a "Telepítési ütemterv" táblákat tartalmazza.

Miután egy új **construction.db** adatbázist hoztunk létre ezekben a táblákban, az adatokat SQL_queries vagy importálás útján lehet hozzáadni ezekhez a táblákhoz, ami lehetővé teszi, hogy később automatikus adatfeldolgozást hozzon létre. Az adatok importálhatók az SQLite adatbázisba CSV fájlokból, Excel táblázatokból vagy exportálhatók más adatbázisokból és adattárakból az API segítségével.

A fenntartható adatmodellezési és hatékony adatbázis-kezelési folyamatok kialakításához a vállalatnak világosan meghatározott stratégiára, valamint a műszaki és üzleti csapatok közötti koordinációra van szüksége. Az eltérő projektek és a több adatforrás esetén gyakran nehéz biztosítani a konziszenciát, a szabványosítást és a minőségellenőrzést minden szinten. Az egyik legfontosabb megoldás lehet egy dedikált adatmodellezési kiválósági központ (CoE) létrehozása a vállalaton belül.

Adatmodellezési kiválósági központ (CoE)

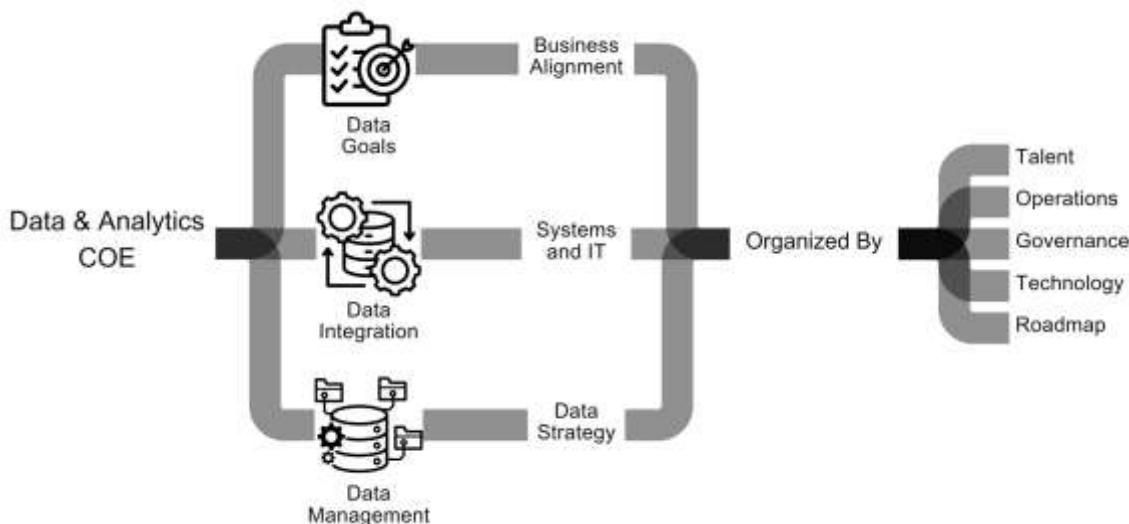
Mivel az adatok az egyik legfontosabb stratégiai eszközökévé váltak, a vállalatoknak nem elég az információkat megfelelően összegyűteni és tárolni - fontos, hogy megtanulják, hogyan kell az adatokat szisztematikusan kezelni. Az osztályozással és adatmodellezéssel foglalkozó kiválósági központ (Centre of Excellence for Classification and Data Modelling (CoE)) egy olyan strukturális egység, amely biztosítja a szervezeten belüli összes adatkezelés következetességét, minőségét és hatékonyságát.

A kiválósági központ (Centre of Excellence, CoE) a szakértői támogatás magja és a vállalat digitális átalakulásának módszertani alapja. Adatvezérelt kultúrát épít ki, és lehetővé teszi a szervezetek számára, hogy olyan folyamatokat alakítsanak ki, amelyek strukturált, validált és reprezentatív adatok alapján hozzák meg döntéseiket, nem pedig intuíció vagy helyi információk alapján.

A kiválósági adatközpont általában a "két pizza" elve szerint működő, több funkciót átfogó csapatokból áll. Ez a Jeff Bezos által javasolt elv azt jelenti, hogy a csapat méretének olyannak kell lennie, hogy két pizzával etethető legyen, azaz legfeljebb 6-10 fővel. Ez a megközelítés segít elkerülni a túlzott bürokráciát és növeli a munka rugalmasságát. A CoE-csapatnak különböző technikai készségekkel rendelkező munkatársakat kell magában foglalnia, az adatalemzéstől és a gépi tanulástól az egyes üzleti területek szakértelmeig. Az adatmérnököknek mély technikai tudásukkal nemcsak a folyamatokat kell optimalizálniuk és az adatokat modellezniük, hanem a rutinfeladatokra fordított idő csökkentésével támogatniuk is kell a kollégákat (4.3-9. ábra).

Ahogy a természetben az ökoszisztemák rugalmasságát a biológiai sokféleség biztosítja, a digitális világban a rugalmasság és az alkalmazkodóképesség az adatok kezelésének sokféle megközelítésével érhető el. Ezt a sokféleséget azonban közös szabályoknak és fogalmaknak kell alátámasztaniuk.

Egy kiválósági központ (CoE) egy erdei ökoszisztéma "éghajlati viszonyaihoz" hasonlítható, amelyek meghatározzák, hogy mely adattípusok fognak virágözni, és melyek fognak automatikusan elvetésre kerülni. A minőségi adatok számára kedvező "klíma" megteremtésével a társközpont elősegíti a legjobb gyakorlatok és módszerek természetes kiválasztását, amelyek aztán a szervezet szabványai lesznek.



Ábra: 4.3-9 Az adat- és elemzési kiválósági központ (CoE) egyesíti az adatkezelés, az integráció és a stratégia kulcsfontosságú szempontjainak szakértelmét.

Az integrációs ciklusok felgyorsítása és a jobb eredmények elérése érdekében az Európa Tanács megfelelő mértékű autonómiát kell biztosítania tagjai számára a döntéshozatalban. Ez különösen fontos egy dinamikus környezetben, ahol a próba és hiba, a folyamatos visszajelzés és a gyakori kiadások jelentős előnyökkel járhatnak. Ez az autonómia azonban csak akkor hatékony, ha a felső vezetés részéről egyértelmű kommunikáció és támogatás áll rendelkezésre. Stratégiai jövőkép és felső szintű koordináció nélkül még a legkompetensebb csapat is akadályokba ütközhet kezdeményezései megvalósításában.

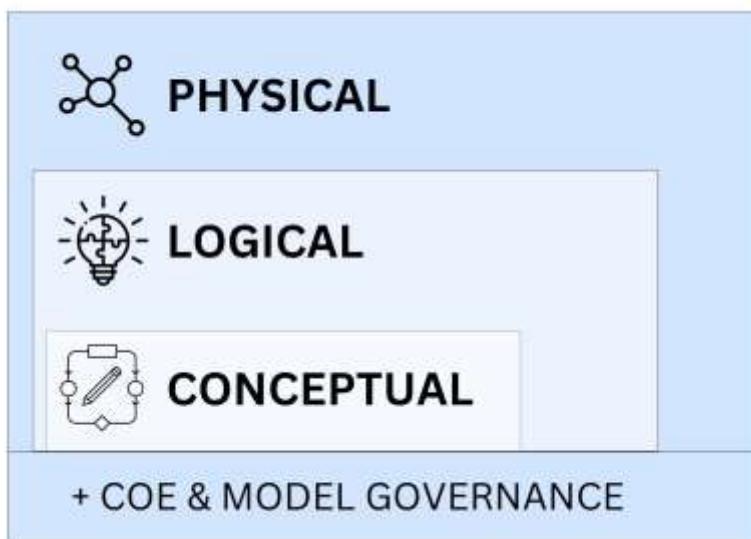
A COE vagy a vállalat felső vezetése felelős azért, hogy az adatmodellezési megközelítés ne csak egy vagy két projektre korlátozódjon, hanem beépüljen a teljes információkezelési és üzleti folyamatirányítási rendszerbe.

A szakértői központ az adatmodellezéssel és az adatkezeléssel kapcsolatos feladatak mellett az adatinfrastruktúra kiépítésére és működtetésére vonatkozó közös szabványok és megközelítések kidolgozásáért is felelős. Emellett elősegíti a szervezeten belül a folyamatos fejlesztés, a folyamatoptimalizálás és az adatok hatékony felhasználásának kultúráját (4.3-10. ábra).

Az adat- és modellkezelés szisztematikus megközelítése az Európa Tanácson belül nagyjából több kulcsfontosságú blokkra osztható:

- **A folyamatok szabványosítása és a modelléletciklus-menedzsment:** a szakpolitikai társulás módszertanokat dolgoz ki és vezet be az adatmodellek létrehozásának és kezelésének egységesítésére. Ez magában foglalja a következőket: strukturális sablonok, minőségellenőrzési módszerek és verzióellenőrzési rendszerek létrehozása az adatok folyamatosságának biztosítása érdekében a munka minden fázisában.

- **Szerepkörök kezelése és felelősségkiosztás:** A CoE meghatározza az adatmodellezési folyamat kulcsszerepeit. A projekt minden résztvevője egyértelműen meghatározott szerepeket és felelősségi köröket kap, ami megkönyíti a csapatmunkát és csökkenti az adatok ellentmondásosságának kockázatát.
- **Minőségellenőrzés és auditálás:** az építési adatok hatékony kezelése megköveteli azok minőségének folyamatos nyomon követését. Az adatok ellenőrzésére, a hibák és a hiányzó attribútumok azonosítására automatizált mechanizmusokat vezetnek be.
- **Metaadat- és információs architektúra-menedzsment:** A CoE felelős a osztályozási és azonosító, elnevezési és entitásleírási szabványok egységes rendszerének létrehozásáért, ami kritikus fontosságú a rendszerek közötti integráció szempontjából.



Ábra 4.3-10 Az adatmodellezés és az adatminőség-menedzsment az egyik fő kihívás a társult vállalkozások számára

A kiválósági központ (CoE) az adatokért nem csupán egy szakértői csoport, hanem egy olyan rendszerszintű mechanizmus, amely új, adatvezérelt kultúrát hoz létre, és biztosítja az adatokkal való munka egységes megközelítését az egész vállalaton belül. A modellezési folyamatoknak a teljes információkezelési rendszerbe való hozzáértő integrálásával, a szabványosítással, az osztályozással és az adatminőség-ellenőrzéssel a CoE segíti a vállalkozásokat abban, hogy folyamatosan javítsák termékeiket és üzleti folyamataikat, gyorsabban reagáljanak a piaci változásokra, és megbízható elemzések alapuló, megalapozott döntéseket hozzanak.

Az ilyen központok különösen hatékonyak, ha a modern DataOps elvekkel kombinálják - a alatt egy olyan lépés, amely biztosítja az adatok folyamatos szállítását, automatizálását és minőségellenőrzését. A DataOps-ról bővebben a 8. részben, a "Modern adatechnológiák az építőiparban" című fejezetben beszélünk.

A következő fejezetekben a stratégiától a gyakorlat felé haladunk - "alakuljunk át" feltételesen egy adatközponttá: több példát is megnézünk arra, hogyan zajlik a feladatak paraméterezése, a követelmények összegyűjtése és az automatikus érvényesítési folyamat.



FEJEZET 4.4.

A KÖVETELMÉNYEK RENDSZEREZÉSE ÉS AZ INFORMÁCIÓK VALIDÁLÁSA

A követelmények összegyűjtése és elemzése: A kommunikáció strukturált adatokká történő átalakítása.

A követelmények összegyűjtése és kezelése az első lépés az adatminőség biztosításához. A digitális eszközök fejlődése ellenére a legtöbb követelményt még mindig strukturálatlan módon fogalmazzák meg: levelek, tárgyalási jegyzőkönyvek, telefonhívások és szóbeli megbeszélések útján. Ez a kommunikációs forma megnehezíti az információk automatizálását, érvényesítését és újrafelhasználását. Ebben a fejezetben azt vizsgáljuk meg, hogyan lehet a szöveges követelményeket formális struktúrákba fordítani, biztosítva ezzel az üzleti követelmények átláthatóságát és rendszerezését.

A Gartner "Data Quality: Best Practices for Accuracy Insights" (Adatminőség: Legjobb gyakorlatok a pontos meglátásokért) című kutatása kiemeli az adatminőség kritikus fontosságát a sikeres adat- és elemzési kezdeményezésekhez [100]. Megjegyzik, hogy a rossz adatminőség évente átlagosan legalább 12,9 millió dollárba kerül a szervezeteknek, és hogy a megbízható, jó minőségű adatok elengedhetetlenek egy adatvezérelt vállalat létrehozásához.

A strukturált követelmények hiánya ahhoz vezet, hogy ugyanazt az elemet (entitást) és annak paramétereit különböző rendszerekben különböző változatokban tárolhatják. Ez nemcsak a folyamatok hatékonyságát csökkenti, hanem időveszteséget, az információk megkettőzését és az adatok felhasználás előtti újból hitelesítésének szükségességét is eredményezi. Ennek eredményeképpen akár egyetlen hiányosság - egy elveszett paraméter vagy egyetlen helytelenül leírt elem - is lassíthatja a döntéshozatalt és az erőforrások nem hatékony felhasználását okozhatja.

*Szög hiján a patkó elveszett.
Patkó hiján a ló elveszett.
Ló hiján elveszett a lovas.
Lovas hiján elveszett az üzenet.
Üzenet hiján elveszett a csata.
Csata hiján elveszett a királyság.
Mindez egy szög hiján a patkóban.*

- Közmondás [101]

Az adatok kitöltésére és tárolására vonatkozó követelmények elemzése és összegyűjtése az összes érdekelt fél azonosításával kezdődik. Ahogyan a közmondás szerint egyetlen szög elvesztése is kritikus következmények láncolatához vezet, az üzleti életben egyetlen érdekelt fél elvesztése, egy figyelmen kívül hagyott követelmény vagy akár egyetlen paraméter elvesztése is jelentősen befolyásolhatja nemcsak az egyes üzleti folyamatokat, hanem a projekt teljes ökoszisztémáját és a szervezet egészét. Ezért kulcsfontosságú, hogy még azokat az elemeket, paramétereket és szerepeket is azonosítsuk, amelyek első pillantásra jelentéktelennek tűnnek, de később kiderülhet, hogy kritikusak az üzleti fenntarthatóság

szempontjából.

Képzeljük el, hogy egy vállalatnak van egy projektje, ahol az ügyfél egy új kéréssel áll elő: "egy további ablakot kell beépíteni az épület északi oldalára". A "az ügyfél kérése egy új ablak hozzáadására a jelenlegi projekthez" kis folyamatba az építész, az ügyfél, a CAD szakember (BIM), az építésvezető, a logisztikai vezető, az ERP - elemző, a minőségellenőrző mérnök, a biztonsági mérnök, az ellenőrzési vezető és az ingatlankezelő vesz részt.

Még egy kis folyamat is több tucat különböző szakembert érinthet. A folyamat minden egyes résztvevőjének ismernie kell a azon szakemberek követelményeit, akikkel adatszinten kapcsolatban állnak.

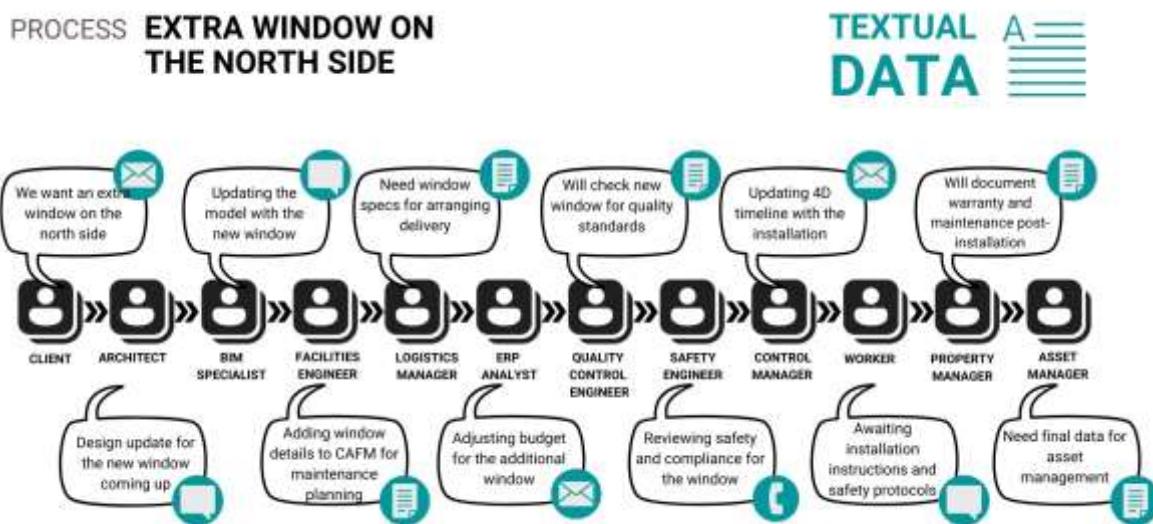
A szöveges szinten (4.4-1. ábra) az ügyfél és a folyamatlánc szakemberei közötti kommunikáció a következőképpen zajlik:

- ⦿ **Ügyfél:** "Úgy döntöttünk, hogy a jobb megvilágítás érdekében az északi oldalon egy további ablakot építünk be. Megvalósítható ez?"
- ⦿ **Építész:** "Persze, átdolgozom a projektet, hogy tartalmazza az új ablakot, és elküldöm a frissített CAD tervezetet (BIM)".
- ⦿ **CAD-szakértő (BIM):** "Új projektet kapott. Frissítem a CAD (BIM) modellt a további ablakkal, és a FEM mérnökkal való egyeztetés után megadom az új ablak pontos helyét és méreteit".
- ⦿ **Építésvezető:** "Új projekt érkezett. Módosítjuk a 4D beépítési dátumait, és tájékoztatjuk az összes érintett alvállalkozót".
- ⦿ **Létesítménymérnök (CAFM):** "Az új ablakon lévő 6D adatokat be fogom vinni a CAFM rendszerbe a jövőbeli létesítménygazdálkodás és karbantartási tervezés céljából."
- ⦿ **Logisztikai vezető:** "Szükségem van az új ablak méreteire és súlyára, hogy megszervezhessem az ablak szállítását a helyszíne".
- ⦿ **ERP -elemző:** "Szükségem van az ERP-rendszerünkben az 5D költségvetés frissítéséhez szükséges hatókör-táblázatokra és a pontos ablaktípusra, hogy az új ablak költségei tükröződjenek a teljes projektbecslésben.".
- ⦿ **Minőségellenőrző mérnök:** "Amint elkészülnek az ablakok specifikációi, meggyőződöm arról, hogy megfelelnek-e a minőségi és anyagszabványainknak."
- ⦿ **Biztonsági mérnök:** "Az új ablak biztonsági szempontjait fogom értékelni, különös tekintettel a 8D rendszer szerinti megfelelőségre és az evakuálásra".
- ⦿ **Ellenőrzési menedzser:** "Az ERP pontos munkaköre alapján frissítjük a 4D idővonalunkat, hogy tükrözze az új ablak beépítését, és tároljuk az új adatokat a projekt tartalomkezelő rendszerében."
- ⦿ **Munkás (telepítő):** "Szüksége van a telepítésre, összeszerelésre és a munka ütemezésére vonatkozó utasításokra. Ezen kívül vannak-e bevezetett különleges biztonsági szabályok, amelyeket be kell tartanom?"
- ⦿ **Ingatlankezelő:** "A telepítést követően dokumentálni fogom a garanciális és karbantartási információkat a hosszú távú kezelés érdekében".
- ⦿ **Vagyonkezelő:** "Berendezésmérnök, kérem, küldje el a végleges adatokat az eszközökötetéshez és az életciklus-kezeléshez".
- ⦿ **Ügyfél:** "Várjon, lehet, hogy sietek, és az ablakra nem lesz szükség. Talán inkább egy erkélyt kellene készítenem".

Az ilyen, gyakran előforduló forgatókönyvek esetében már egy apró változás is láncreakciót vált ki több

rendszer és szerepkör között. Ebben az esetben a kezdeti fázisában szinte minden kommunikáció szöveges formában történik: e-mailek, csevegések, tárgyalási jegyzőkönyvek (4.4-1. ábra).

Egy ilyen szöveges kommunikációs rendszerben egy építési projekt esetében nagyon fontos az összes adatcsere-művelet és minden meghozott döntés jogi megerősítésének és rögzítésének rendszere. Ez biztosítja, hogy minden egyes döntés, utasítás vagy módosítás jogilag érvényes és nyomon követhető legyen, csökkentve a jövőbeli "félreértesek" kockázatát

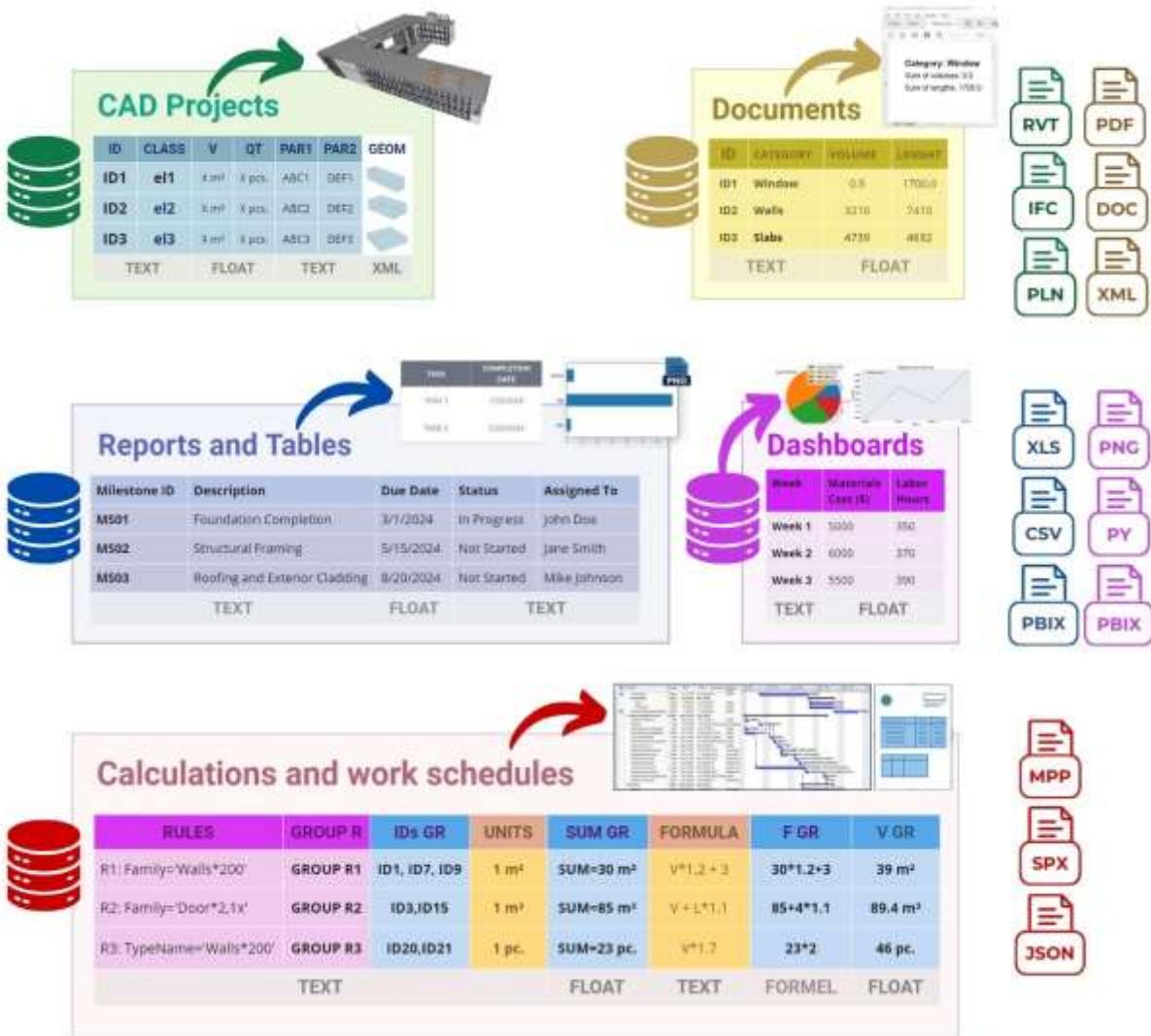


4.4-1 A projekt kezdeti szakaszában az ügyfél és a vállalkozó közötti kommunikáció gyakran tartalmaz több formátumú szöveges adatokat.

A jogi ellenőrzés és a döntések érvényesítésének hiánya az építési projekt vonatkozó rendszereiben komoly problémákhoz vezethet minden érintett számára. minden megfelelő dokumentáció és érvényesítés nélkül hozott döntés, utasítás vagy módosítás vitákhoz (és pereskedéshez) vezethet.

A szöveges kommunikációban hozott valamennyi döntés jogi konszolidációja csak nagyszámú aláírt dokumentummal biztosítható, ami a vezetőség vállára hárul, amely köteles minden tranzakciót rögzíteni. Ennek eredményeképpen, ha minden résztvevőnek minden művelethez dokumentumokat kell aláírnia, a rendszer elveszíti rugalmasságát, és bürokratikus útvesztővé válik. A tranzakciók visszaigazolásának hiánya nemcsak késlelteti a projekt megvalósítását, hanem pénzügyi veszteségekhez és a résztvevők közötti kapcsolatok megromlásához vezethet, egészen a jogi problémáig.

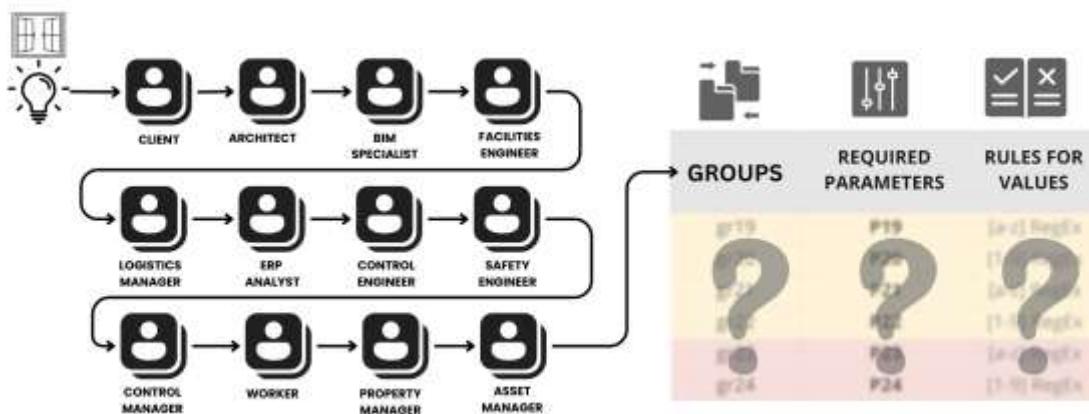
Egy ilyen tranzakció jóváhagyási folyamata, amely általában szöveges megbeszélésekkel kezdődik, a következő szakaszokban fokozatosan több formátumú dokumentumcserévé alakul át (4.4-2. ábra), ami jelentősen bonyolítja a korábban csak szöveges úton zajló kommunikációt. Egyértelműen meghatározott követelmények nélkül az ilyen, több formátumú adatokkal és nagyszámú szöveges követelménnel teli folyamatok automatizálása szinte lehetetlenné válik.



4.4-2. ábra Az építőipari vállalat minden egyes rendszere a jogilag releváns dokumentumok forrásaként szolgál, különböző formátumokban.

A szöveges kommunikáció megköveteli, hogy minden szakember vagy megismerje a teljes levelezést, vagy rendszeresen részt vegyen az összes találkozón, hogy megértse a projekt aktuális helyzetét.

E korlátozás leküzdése érdekében át kell térti a szöveges kommunikációról a strukturált követelménymodellre. Ez csak szisztematikus elemzéssel, a folyamatok vizualizálásával és a kölcsönhatások folyamatábrák és adatmodellek formájában történő leírásával lehetséges (4.4-3. ábra). Az adatmodellezéshez hasonlóan (4.3-7. ábra) a kontextus-ötlet szintről a résztvevők által használt rendszerek és eszközök, valamint a közöttük lévő kapcsolatok hozzáadásával jutottunk el a koncepcionális szintre.



Ábra: 4.4-3 Ahhoz, hogy megtanuljuk, hogyan kell kezelni és automatizálni a validálási folyamatot, szükséges a folyamatok vizualizálása és a követelmények strukturálása.

A követelmények és kapcsolatok rendszerezésének első lépése az összes kapcsolat és kapcsolat vizualizálása fogalmi folyamatábrák segítségével. A fogalmi szintű nemcsak megkönnyíti a folyamat valamennyi résztvevője számára a teljes folyamatlánc megértését, hanem világosan megmutatja azt is, hogy miért és kinek van szüksége az adatokra (és követelményekre) az egyes folyamatlépésekben.

Folyamatábrák és a fogalmi keretek hatékonyiséga

Az adatkezelés hagyományos és modern megközelítései közötti szakadék áthidalása érdekében a vállalatoknak tudatosan át kell térniük a töredékes szöveges leírásokról a strukturált folyamatreprezentációkra. Az adatok fejlődése - az agyagtáblától a digitális ökoszisztemáig - új gondolkodási eszközöket igényel. Az egyik ilyen eszköz pedig a folyamatábrák segítségével történő fogalmi modellezés. A visuális diagramok - folyamatábrák, folyamatábrák, interakciós diagramok - készítése lehetővé teszi a projekt résztvevői számára, hogy felismerjék, hogyan hatnak cselekedeteik és döntéseiük a teljes döntéshozatali rendszerre.

Ha a folyamatok nemcsak az adatok tárolását, hanem azok elemzését vagy automatizálását is igénylik, akkor el kell kezdenie foglalkozni a koncepcionális és visuális követelményréteg létrehozásának téma jával.

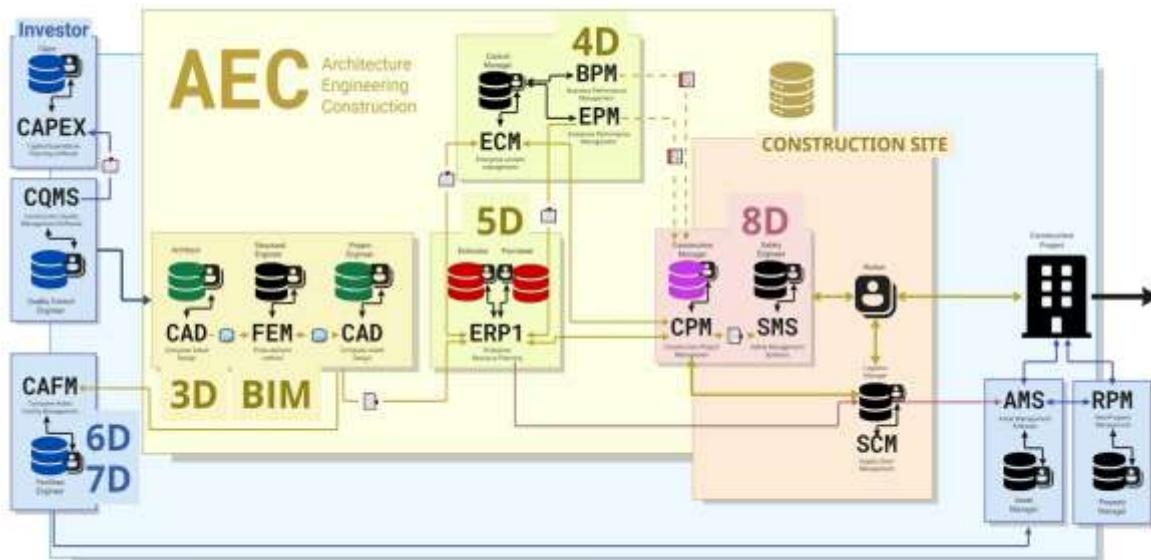
Példánkban (4.4-1. ábra) az egyes szakemberek egy kis csapat tagjai lehetnek, de egy nagyobb, akár egy tucatnyi szakembert is magában foglaló részlegnek is, amelyet egy főigazgató irányít. minden részleg egy speciális alkalmazási adatbázist használ (1.2-4. ábra, pl. ERP, CAD, MEP, CDE, ECM, CPM stb.), amelyet rendszeresen frissítnek a dokumentumok létrehozásához, a döntések jogi státuszának rögzítéséhez és a folyamatok irányításához szükséges beérkező információkkal.

A tranzakciós folyamat hasonlít az ókorai menedzserek 4000 évvel ezelőtti munkájához, amikor agyagtáblákat és papiruszt használtak a döntések jogi megerősítésére. A különbség a modern rendszerek és az agyag és papír elődeik között az, hogy a modern módszerek ezen felül magukban foglalják a szöveges információk digitális formába történő átalakításának folyamatát, hogy azokat más rendszerekben és eszközökben további automatikus feldolgozásra lehessen használni.

A folyamat vizualizálása koncepcionális folyamatábrák formájában segít leírni az egyes lépéseket és a különböző szerepek közötti kölcsönhatásokat, így a bonyolult munkafolyamatok áttekinthetővé és egyszerűvé válnak.

A folyamatok vizualizálása biztosítja, hogy a folyamat logikája átlátható és elérhető legyen a csapat minden tagja számára.

Ugyanaz a kommunikációs folyamat, amely egy ablak hozzáadására irányul egy projekthez, és amelyet szöveg, üzenetek (4.4-1. ábra) és blokkdiagram formájában leírtunk, hasonló az adatmodellezésről szóló fejezetben tárgyalt koncepcionális modellhez (4.4-4. ábra).



Ábra 4.4-4 A koncepcionális diagram a projekt résztvevőit egy adatbázis felhasználóiként mutatja be, ahol a lekérdezések összekapcsolják a különböző rendszereket.

Bár a koncepcionális diagramok fontos lépést jelentenek, sok vállalat csak erre a szintre szorítkozik, mivel úgy gondolja, hogy a folyamatok megértéséhez elegendő egy vizuális diagram. Ez a kezelhetőség illúzióját kelti: a vezetők egy ilyen folyamatábrán könnyebben érzékelik az összképet, és könnyebben látták a résztvevők és a szakaszok közötti kapcsolatokat. Az ilyen sémák azonban nem adnak világos képet arról, hogy az egyes résztvevők esetében milyen adatokra van szükség, milyen formátumban kell azokat továbbítani, és milyen paraméterek és attribútumok kötelezők az automatizálás megvalósításához. A fogalmi folyamatábra inkább egy útvonalterképhez hasonlít: jelzi, hogy ki kivel lép kölcsönhatásba, de nem mutatja meg, hogy ezek során mi kerül átadásra.

Még ha egy folyamatot koncepcionális szinten, folyamatábrák segítségével részletesen le is írunk, ez nem garantálja annak hatékonyságát. A vizualizáció gyakran egyszerűsíti a vezetők munkáját, lehetővé téve számukra, hogy lépésről lépésre történő jelentéssel könnyebben nyomon követhessék a folyamatot. Az adatbázis-mérnökök számára azonban előfordulhat, hogy a koncepcionális ábrázolás nem elég világos, és nem biztos, hogy egyértelműen megértik, hogyan kell a folyamatot paraméter- és követelményszinten megvalósítani.

Ahogy haladunk az összetettebb adatökoszisztémák felé, a koncepcionális és vizuális eszközök kezdeti bevezetése kritikus jelentőségűvé válik annak biztosítása érdekében, hogy az adatfolyamatok ne csak hatékonyak legyenek, hanem összhangban legyenek a szervezet stratégiai céljaival is. Ahhoz, hogy ezt az ablak hozzáadási folyamatot (4.4-1. ábra) teljes mértékben lefordítsuk az adatkövetelmények szintjére, egy szinttel mélyebbre kell mennünk, és a folyamat fogalmi vizualizációját le kell fordítanunk az adatok, a szükséges attribútumok és azok határértékeinek logikai és fizikai szintjére.

Strukturált követelmények és RegEx reguláris kifejezések

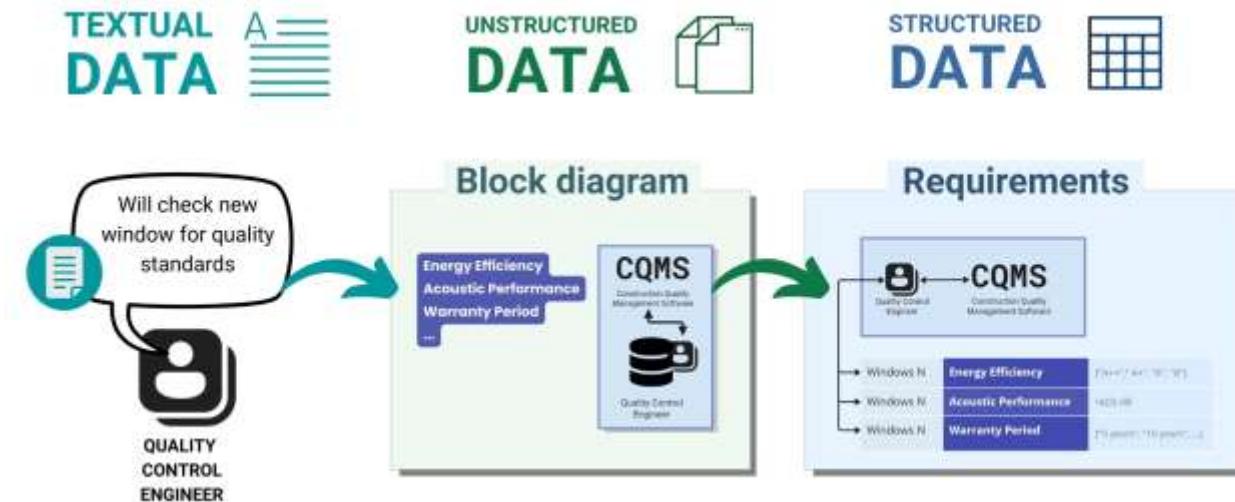
A vállalatoknál keletkező adatok akár 80%-a strukturálatlan vagy félig strukturált formátumú [52] - szöveg, dokumentumok, levelek, PDF - fájlok, beszélgetések. Az ilyen adatokat (4.4-1. ábra) nehéz elemezni, ellenőrizni, a rendszerek között átvinni és az automatizálásban felhasználni.

A kezelhetőség, az átláthatóság és az automatikus érvényesítés biztosítása érdekében a szöveges és félig strukturált követelményeket jól meghatározott, strukturált formátumokra kell lefordítani. A strukturálási folyamat nemcsak az adatokra vonatkozik (amelyeket a könyv ezen részének első fejezeteiben részletesen tárgyaltunk), hanem magukra a követelményekre is, amelyeket a projekt résztvevői általában szabad szöveges formában fogalmaznak meg a projekt életciklusa során, gyakran anélkül, hogy arra gondolnának, hogy ezek a folyamatok automatizálhatók.

Ahogyan az adatokat strukturálatlan szöveges formából strukturált formába alakítottuk át, a követelmények munkafolyamatában a szöveges követelményeket strukturált "logikai és fizikai réteg" formátumba alakítjuk át.

Az ablak hozzáadására vonatkozó példa (4.4-1. ábra) részeként a következő lépés az adatigények táblázatos formában történő leírása. A projekt résztvevői által használt minden egyes rendszerre vonatkozó információt a kulcsfontosságú attribútumok és azok határértékeinek megadásával strukturáljuk

Vegyük például egy ilyen rendszert (4.4-5. ábra) - az építési minőségirányítási rendszert (CQMS), amelyet az ügyfél oldalán a minőségellenőrző mérnök használ. Ennek segítségével ellenőrzi, hogy a projekt egy új eleme - jelen esetben az "új ablak" - megfelel-e a megállapított szabványoknak és követelményeknek.



4.4-5 A szöveges követelmények táblázatos formátumba történő átalakítása az entitás attribútumainak leírásával megkönyíti a megértést más szakemberek számára.

Példaként tekintsünk meg néhány fontos követelményt a CQMS -rendszer "ablakrendszer" típusú entitások attribútumaira vonatkozóan (4.4-6. ábra): energiahatékonyság, akusztikai teljesítmény és garanciaidő. Mindegyik kategória tartalmaz bizonyos szabványokat és előírásokat, amelyeket figyelembe kell venni az ablakrendszer tervezésekor és beépítésekor.



Ábra 4.4-6 A minőségellenőrző mérnöknek az új Ablak típusú elemeket az energiahatékonyság, a hangszigetelés és a garanciális előírások szempontjából kell megvizsgálnia.

A minőségellenőrző mérnök által táblázat formájában megadott adatkövetelmények például a következő határértékekkel rendelkeznek:

- **Az ablakok energiahatékonysági osztálya** a legmagasabb hatékonyságot jelölő "A++"-tól a minimálisan elfogadható szintnek tekintett "B"-ig terjed, és ezeket az osztályokat az elfogadható értékek listája ["A++", "A+", "A", "A", "A", "A", "A", "B"] jelöli.
- **Az ablakok hangszigetelését**, amelyet decibelben mérnek, és amely megmutatja, hogy mennyire

képesek csökkenteni az utcai zajokat, a \d{2}dB szabályos kifejezéssel határozzák meg.

- Az "Ablak típusa" entitás "**Garanciaidőszak**" attribútuma öt évnél kezdődik, ezt az időtartamot határozza meg a termék kiválasztásakor megengedett minimális időtartamként; a garanciaidő értékei, mint például ["5 év", "10 év" stb.] vagy a ">5 (év)" logikai feltétel is megadható.

Az összegyűjtött követelmények szerint a megállapított attribútumokon belül a "B" alatti, például "C" vagy "D" osztályzatú új ablakkategória vagy osztályelemek nem felelnek meg az energiahatékonysági vizsgálaton. Az ablakok hangszerelését a minőségbiztosítási mérnöknek küldött adatokban vagy dokumentumokban kétjegyű számmal kell jelölni, amelyet a "dB" utótag követ, például "35 dB" vagy "40 dB", és az ezen a formátumon kívüli értékeket, például "9 D B" vagy "100 decibel" nem fogadják el (mivel ezek nem mennek át a RegEx karakterláncok mintázóján). A garanciaidőnek legalább "5 ével" kell kezdődnie, és a rövidebb garanciaidővel rendelkező ablakok, mint például "3 év" vagy "4 év" nem fognak megfelelni a minőségügyi mérnök által a táblázat formátumában leírt követelményeknek.

Az ilyen attribútum-paraméter értékek ellenőrzésére a követelményekből származó határértékekkel szemben az érvényesítési folyamat során vagy a megengedett értékek listáját (["A", "B", "C"]), vagy szótárat (["A": "H1", "H2", "B": "W1", "W2"]), logikai műveletek (pl. ">", "<", "<=", ">=" "==") numerikus értékek esetén) és reguláris kifejezések (karakterlánc- és szövegértékek esetén, mint például az "Akusztikai teljesítmény" attribútumban). A reguláris kifejezések rendkívül fontos eszköz a karakterláncértékekkel való munka során.

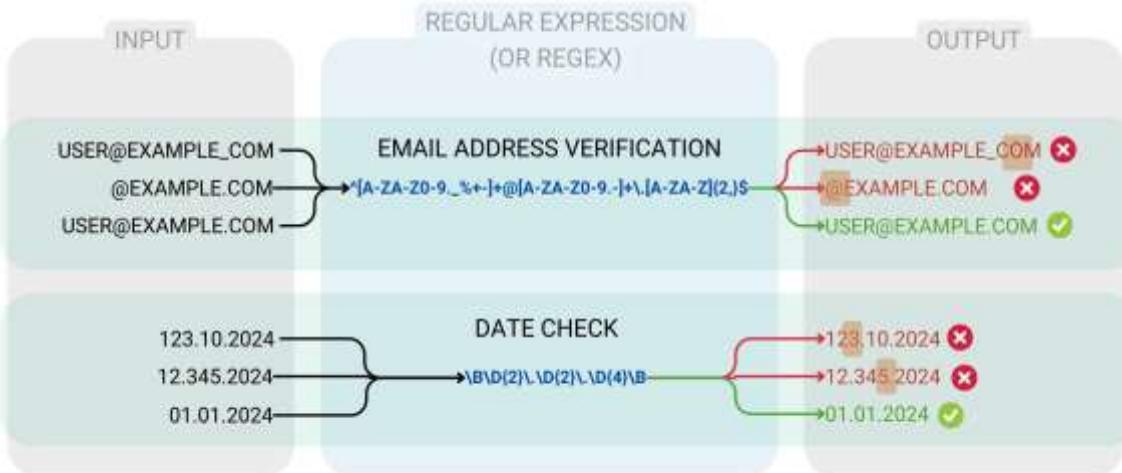
A reguláris kifejezéseket (RegEx) a programozási nyelvekben, többek között a Pythonban (Re könyvtár) használják karakterláncok keresésére és módosítására. A Regex olyan, mint egy detektív a karakterláncok világában, amely képes a szövegen lévő szövegmintákat precízen azonosítani.

A reguláris kifejezésekben a betűket közvetlenül a megfelelő ábécé karakterekkel írjuk le, míg a számok a \d speciális karakterrel ábrázolhatók, amely 0 és 9 közötti bármely számjegynek megfelel. A szöglletes zárójelek a betűk vagy számjegyek tartományának jelölésére szolgálnak, pl. [a-z] a latin ábécé bármely kisbetűs betűjével, vagy [0-9], ami a \d-nek felel meg. A nem számjegyek és a nem betűjelek esetében \D és \W használatos.

Népszerű RegEx felhasználási esetek (4.4-7. ábra):

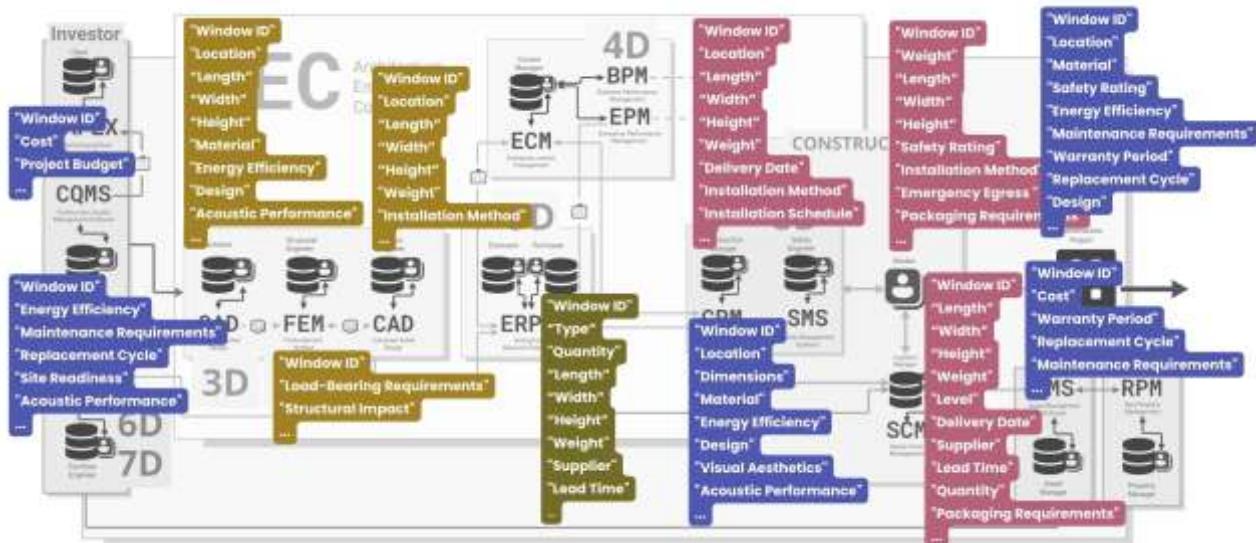
- **e-mail cím ellenőrzése:** Ha ellenőrizni szeretné, hogy egy karakterlánc érvényes e-mail cím-e, használhatja a "^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}\$" sablont.
- **Dátum kinyerése:** A "\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b" sablon használható a dátum kinyerésére szövegből DD.MM.YYYY formátumban.
- **telefonszámok ellenőrzése:** a +49(000)000-0000 formátumú telefonszámok ellenőrzéséhez a minta a következőképpen néz ki: "\+\d{2}\(\d{3}\)\d{3}-\d{4}".

A minőségbiztosítási mérnök követelményeinek az attribútumok és határértékeik formátumára történő lefordításával (4.4-6. ábra) az eredeti szöveges formátumból (beszélgetések, levelek és szabályozási dokumentumok) egy szervezett és strukturált táblázattá alakítottuk át őket, így lehetővé téve a beérkező adatok (pl. az Ablak kategória új elemei) automatikus ellenőrzését és elemzését. A követelmények megléte lehetővé teszi, hogy a nem ellenőrzött adatok automatikusan elvetésre kerüljenek, és az ellenőrzött adatok automatikusan továbbításra kerüljenek a rendszerekbe további feldolgozás céljából.



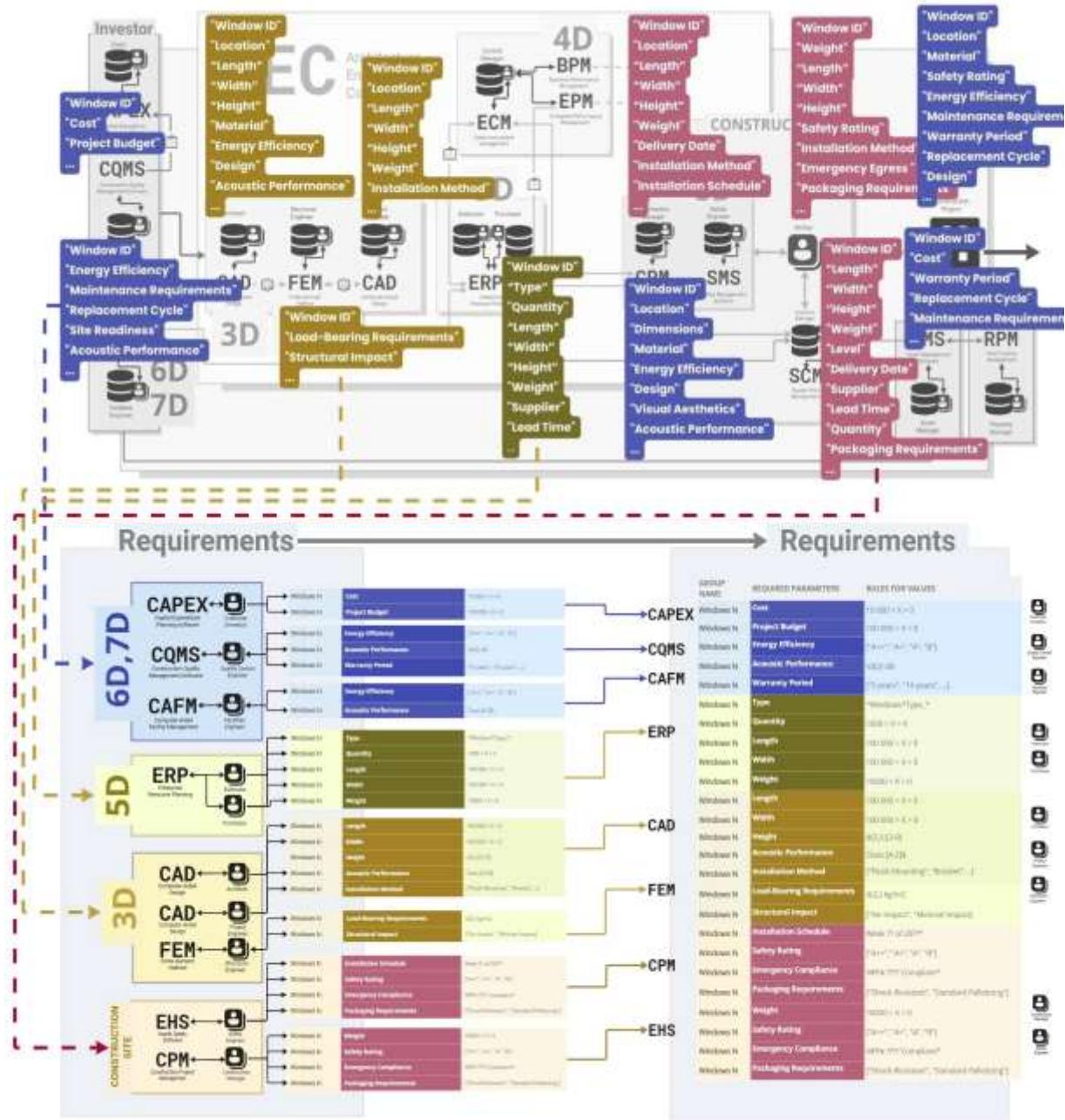
4.4-7 ábra A reguláris kifejezések használata rendkívül fontos eszköz a szöveges adatok érvényesítési folyamatában.

Most, a követelményekkel való munka fogalmi szintjéről a logikai szintre lépve, az új ablak telepítésének folyamatában (4.4-4. ábra) az összes szakember összes követelményét () attribútum formátumú rendezett listává alakítjuk, és ezeket a listákat a szükséges attribútumokkal hozzáadjuk az egyes szakemberek folyamatábrájához (4.4-8. ábra).



4.4-8. ábra A logikai folyamatok szintjén az egyes szakemberek által kezelt attribútumok hozzáadódnak a megfelelő rendszerekhez.

Azáltal, hogy az összes attribútumot egyetlen közös folyamat-táblázatba foglaljuk, a korábban szövegként és párbeszédként a fogalmi szinten (4.4-1. ábra) megjelenített információkat a fizikai szintű táblázatok strukturált és rendszerezett formájába (4.4-9. ábra) alakítjuk át.



4.4-9 ábra A strukturálatlan szakmai párbeszéd strukturált táblázatokká történő átalakítása segít megérteni a követelményeit fizikai szinten.

Most az adatigényeket kell közölni az egyes rendszerekhez információkat létrehozó szakemberekkel. Ha például egy CAD-adatbázisban dolgozik, az elemek modellezésének megkezdése előtt össze kell gyűjtenie az összes szükséges paramétert az adatok végfelhasználási forgatókönyvei alapján. Ez általában az üzemeltetési fázissal kezdődik, ezt követi az építkezés, a logisztikai osztály, a becslési osztály, a statikai számítások osztálya és így tovább. Csak miután figyelembe vette minden területek követelményeit, kezdheti el az adatok létrehozását - az összegyűjtött paraméterek alapján. Ez lehetővé teszi, hogy a jövőben automatizálja az adatok ellenőrzését és továbbítását a lánc mentén.

Ha az új adatok megfelelnek a követelményeknek, automatikusan integrálódnak a vállalat adatökoszisztémájába, és közvetlenül azokhoz a felhasználókhoz és rendszerekhez jutnak el, akiknek szánták őket. Az adatok ellenőrzése az attribútumok és azok értékei alapján biztosítja, hogy az információ megfelel az előírt minőségi szabványoknak, és készen áll a vállalati forgatókönyvek alkalmazásához.

Az adatigényeket már meghatározták, és most, mielőtt az ellenőrzés megkezdődhetne, az ellenőrizendő adatokat létre kell hozni, be kell szerezni vagy össze kell gyűjteni, vagy az adatbázisokban lévő információk aktuális állapotát rögzíteni kell, hogy az ellenőrzés során felhasználhatók legyenek.

Adatgyűjtés az ellenőrzési folyamathoz

Mielőtt a validálás megkezdődne, fontos biztosítani, hogy az adatok a validálási folyamathoz megfelelő formában álljanak rendelkezésre. Ez nem egyszerűen azt jelenti, hogy az információ rendelkezésre áll, hanem azt, hogy elő kell készíteni: az adatokat össze kell gyűjteni és strukturálatlan, lazán strukturált, szöveges és geometriai formátumokból strukturált formába kell alakítani. Ezt a folyamatot az előző fejezetek részletesen ismertetik, ahol a különböző adattípusok átalakításának módszereit tárgyalják. Az összes átalakítás eredményeként a beérkező adatok nyitott strukturált táblázatok formájában jelennek meg (4.1-2. ábra, 4.1-9. ábra, 4.1-13. ábra).

A követelmények és a szükséges paramétereket és határértékeket tartalmazó strukturált táblázatok (4.4-9. ábra) birtokában megkezdhetjük az adatok érvényesítését - akár egyetlen automatizált folyamatként (Pipeline), akár az egyes beérkező dokumentumok lépésről lépésre történő érvényesítésével.

Az ellenőrzés elindításához vagy egy új fájlt kell bemenetként fogadni, vagy rögzíteni kell az adatok jelenlegi állapotát - pillanatfelvételt kell készíteni, vagy exportálni kell a jelenlegi és a beérkező adatokat, vagy kapcsolatot kell létesíteni egy külső vagy belső adatbázissal. A vizsgált példában egy ilyen pillanatfelvétel úgy jön létre, hogy a CAD-adatokat automatikusan átalakítják egy strukturált formátumba, amelyet mondjuk 2024. március 29-én, pénteken 23:00:00 órakor rögzítenek, miután az összes tervező hazament.



4.4-10. ábra CAD-adatbázis pillanatkép (BIM), amely a projektmodell aktuális verziójában az "Ablak" osztályú új entitás aktuális attribútumadatait mutatja.

A "CAD-adatok (BIM) strukturált formába történő fordítása" című fejezetben tárgyalott reverse engineering eszközöknek köszönhetően a különböző CAD (BIM) eszközökből és szerkesztőkből származó információk külön táblázatokba rendezhetők (4.4-11. ábra) vagy a projekt különböző szakaszait egyesítő közös

táblázattá kombinálhatók (9.1-10. ábra).

Egy ilyen tábla - adatbázis megjeleníti az ablakok és ajtók egyedi azonosítóit (ID attribútum), típusneveit (TypeName), méreteit (Width, Length), anyagait (Material), valamint az energetikai és akusztikai teljesítménymutatókat és egyéb jellemzőket. Egy ilyen, a (BIM) CAD-programban kitöltött táblázatot a tervezőmérnök a különböző részlegekből és dokumentumokból gyűjti össze, így alkotva a projekt információs modelljét.



4.4-11. ábra A CAD rendszerekből származó strukturált adatok lehetnek kétdimenziós táblázatok, amelyek oszlopai az elemek attribútumait jelölik.

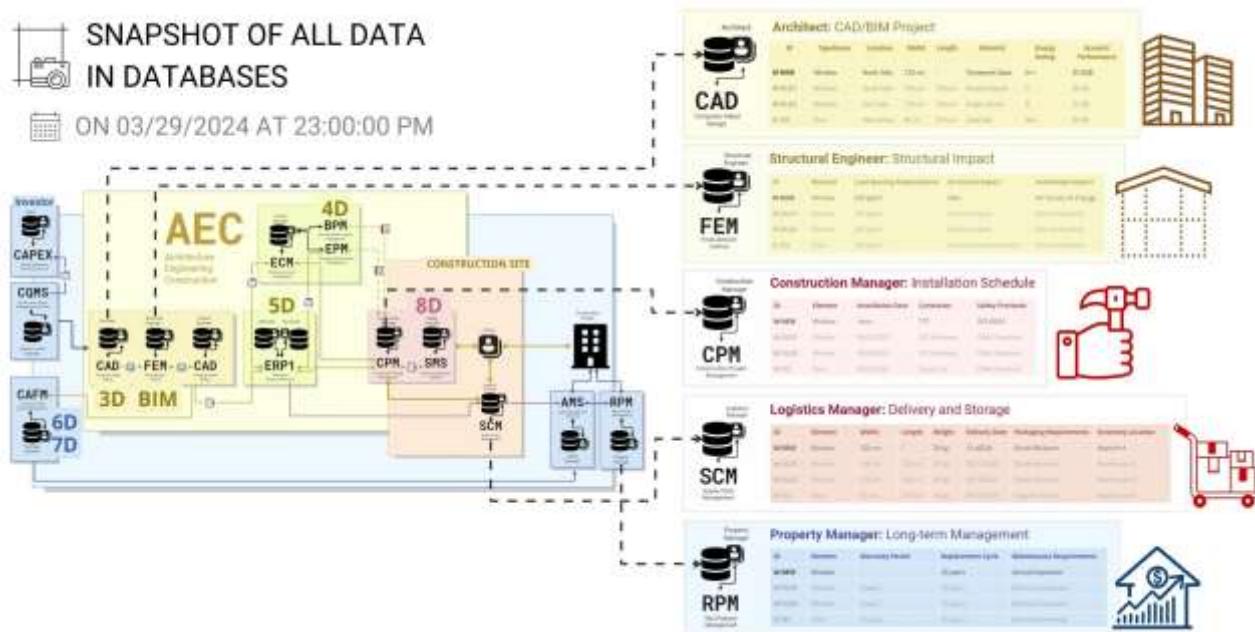
A valódi CAD (BIM) projektek több tíz- vagy százezer elemet tartalmaznak (9.1-10. ábra). A CAD-formátumokon belüli elemek automatikusan típus és kategória szerint kategorizálódnak - az ablakoktól és ajtóktól a födémekig, födémekig és falakig. Egyedi azonosítók (pl. natív azonosító, amelyet a CAD-megoldás automatikusan beállít) vagy típusattribútumok (típusnév, típus, család) lehetővé teszik ugyanazon objektum nyomon követését a különböző rendszerekben. Például egy épület északi falán lévő új ablakot a szervezet összes releváns rendszerében egyetlen "W-NEW" azonosítóval lehet egyértelműen azonosítani.

Míg az entitások neveinek és azonosítóinak minden rendszerben konzisztensnek kell lenniük, az ezekhez az entitásokhoz kapcsolódó attribútumok és értékek a felhasználási kontextustól függően jelentősen eltérhetnek. Az építészek, a statikusok, az építőipari, logisztikai és ingatlanfenntartási szakemberek minden másképp érzékelik ugyanazokat az elemeket.

Mindegyikük a saját osztályozóikra, szabványaikra és céljaikra támaszkodik: egyesek tisztán esztétikai szempontból vizsgálnak egy ablakot, értékelve annak formáját és arányait, míg mások mérnöki vagy üzemeltetési szempontból, elemezve a hővezető képességet, a beépítési módszert, a súlyt vagy a karbantartási követelményeket. Ezért az adatok modellezésekor és az elemek leírásakor fontos figyelembe venni a felhasználás sokoldalúságát és biztosítani az adatok konzisztenciáját, miközben figyelembe kell venni az iparági sajátosságokat.

A vállalat folyamataiban betöltött minden egyes szerepkörhöz saját felhasználói felülettel rendelkező speciális adatbázisok tartoznak - a tervezéstől és számításuktól kezdve a logisztikán át a telepítésig és az épület üzemeltetéséig (4.4-12. ábra). minden ilyen rendszert egy-egy szakemberekből álló szakmai csapat kezel egy speciális felhasználói felületen vagy adatbázis-lekérdezésekben keresztül, ahol a lánc végén a bevitt értékek alapján hozott döntések összessége mögött a rendszer- vagy osztályvezető áll, aki a más

rendszerket kiszolgáló partnerei felé a bevitt adatok jogi érvényességéért és minőségéért felel.



4.4-12. ábra Ugyanaz az entitás különböző rendszerekben ugyanazzal az azonosítóval rendelkezik, de különböző attribútumokkal, amelyek csak az adott rendszerben fontosak.

Miután megszerveztük a strukturált követelmények és adatok gyűjtését logikai és fizikai szinten, már csak egy olyan folyamatot kell létrehoznunk, amely automatikusan ellenőrzi a különböző beérkező dokumentumokból és rendszerekből származó adatokat a korábban összegyűjtött követelményekkel.

A adatok ellenőrzése és az ellenőrzés eredményei

A rendszerbe kerülő minden új adatot - legyenek azok dokumentumok, táblázatok vagy adatbázis-bejegyzések az ügyfélről, az építészről, a mérnökről, a művezetőről, a logisztikusról vagy az ingatlankezelőről - a korábban megfogalmazott követelményekkel szemben validálni kell (4.4-9. ábra). Az érvényesítési folyamat kritikus jelentőségű: az adatokban előforduló hibák hibás számításokhoz, az ütemterv késedelméhez, sőt pénzügyi veszteségekhez is vezethetnek. Az ilyen kockázatok minimalizálása érdekében szisztematikus és megismételhető, iteratív adatérvényesítési eljárást kell szervezni.

A rendszerbe érkező új - strukturálatlan, szöveges vagy geometriai - adatok érvényesítéséhez azokat lazán strukturált vagy strukturált formátumba kell konvertálni. Az érvényesítési folyamatnak ezután az adatokat az előírt attribútumok és megengedett értékeik teljes listájával kell összevetnie.

A különböző típusú adatok: szöveg, képek, PDF -dokumentumok és vegyes CAD (BIM) adatok strukturált formába történő konvertálását részletesen az "Adatok strukturált formába történő konvertálása" című fejezetben tárgyalunk.

Erre példa egy CAD (BIM) projektből származó táblázat (4.4-11. ábra). Ez félígy strukturált geometriai adatokat és strukturált attribútuminformációkat tartalmaz a projekt entitásokról (3.1-14. ábra) - pl. egy elem az "ablakok" osztályból.

A validálás elvégzéséhez az attribútumértékeket (4.4-11. ábra) összehasonlítjuk a szakértők által követelmény formájában meghatározott referencia határértékekkel (4.4-9. ábra). A végső összehasonlító táblázat (4.4-13. ábra) világossá teszi, hogy mely értékek elfogadhatóak, és melyeket kell korrigálni, mielőtt az adatok CAD-alkalmazásokon kívül is felhasználhatók lennének (BIM).



4.4-13 ábra A végső érvényesítési táblázat kiemeli a "Windows" osztályú új entitás azon attribútumértékeit, amelyekre érdemes odafigyelni.

Egy hasonló megoldás megvalósítása a Pandas könyvtár segítségével, amelyet korábban a "Pandas: Egy nélkülvilágos eszköz az adatelemzéshez" című fejezetben ismertetünk, egy CAD fájlból (BIM) (RVT, IFC, DWG, NWS, DGN) kinyert táblázatos fájlból (4.4-11. ábra) származó adatokat fogunk validálni egy másik táblázatos követelményfájlból (4.4-9. ábra) származó követelmények segítségével.

A kód megszerzéséhez le kell írnunk az LLM kérdőívben, hogy be kell töltenünk a adatokat a **raw_data.xlsx** fájlból (a CAD-adathábazis (BIM) teljes adatkészlete), **ellenőrizni** kell őket, és az eredményt egy új **checked_data.xlsx** fájlba kell menteni (4.4-13. ábra).

- ❷ A kódot az LLM segítségével a Pandas könyvtár megemlíttése nélkül kapjuk meg:

Írunk kódot a raw_data.xlsx fájlból származó táblázat ellenőrzésére és érvényesítésére a következő érvényesítési szabályok alapján: a "Width" és a "Length" oszlopok értékei nagyobbak nullánál, az "Energy Rating" szerepel az ['A++', 'A+', 'A', 'A', 'B'] listában, az "Acoustic Performance" pedig egy később megadandó változó - az utolsó érvényesítési oszlop hozzáadásával, és mentsük a végleges táblázatot egy új Excel fájlba: checked_data.xlsx. ↴

- Az LLM válasza egy rövid Python-kód példát ír le, amelyet a következő kérésekkel lehet finomítani és bővíteni :



```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

4.4-14. ábra Az LLM- modell által generált kód ellenőri a konvertált CAD (BIM) tervet a határértékek formájában megadott attribútumkövetelményekkel szemben.

Az LLM nyelvi modell által generált kód, bármely népszerű IDE-ben vagy online eszközben használható: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev pluginnal, Thonny, Wing IDE, IntelliJ IDEA Python pluginnal, JupyterLab vagy népszerű online eszközökkel Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

A kód végrehajtása (4.4-14. ábra) azt mutatja, hogy a (BIM) CAD-adatbázisból származó W-OLD1, W-OLD2, D-122 (és más elemek) "entitásselemek" megfelelnek az attribútumkövetelményeknek: a szélesség és a hossz nagyobb, mint nulla, és az energiahatékonysági osztály az "A++", "A", "B", "C" listában szereplő értékek egyike (4.4-15. ábra).

A W-NEW elem, amelyre szükségünk van, és amelyet nemrég adtunk hozzá, és amely az északi oldalon az új "Ablak" elemosztályért felelős, nem felel meg a követelményeknek (a "Requrements Met" attribútum), mivel a hossza nulla (a "0.0" értéket a "Width'>0" szabályunk elfogadhatatlannak tekinti), és nem ad meg energiahatékonysági osztályt.


CHECKED_DATA.XLSX (CSV)
VERIFIED DATA

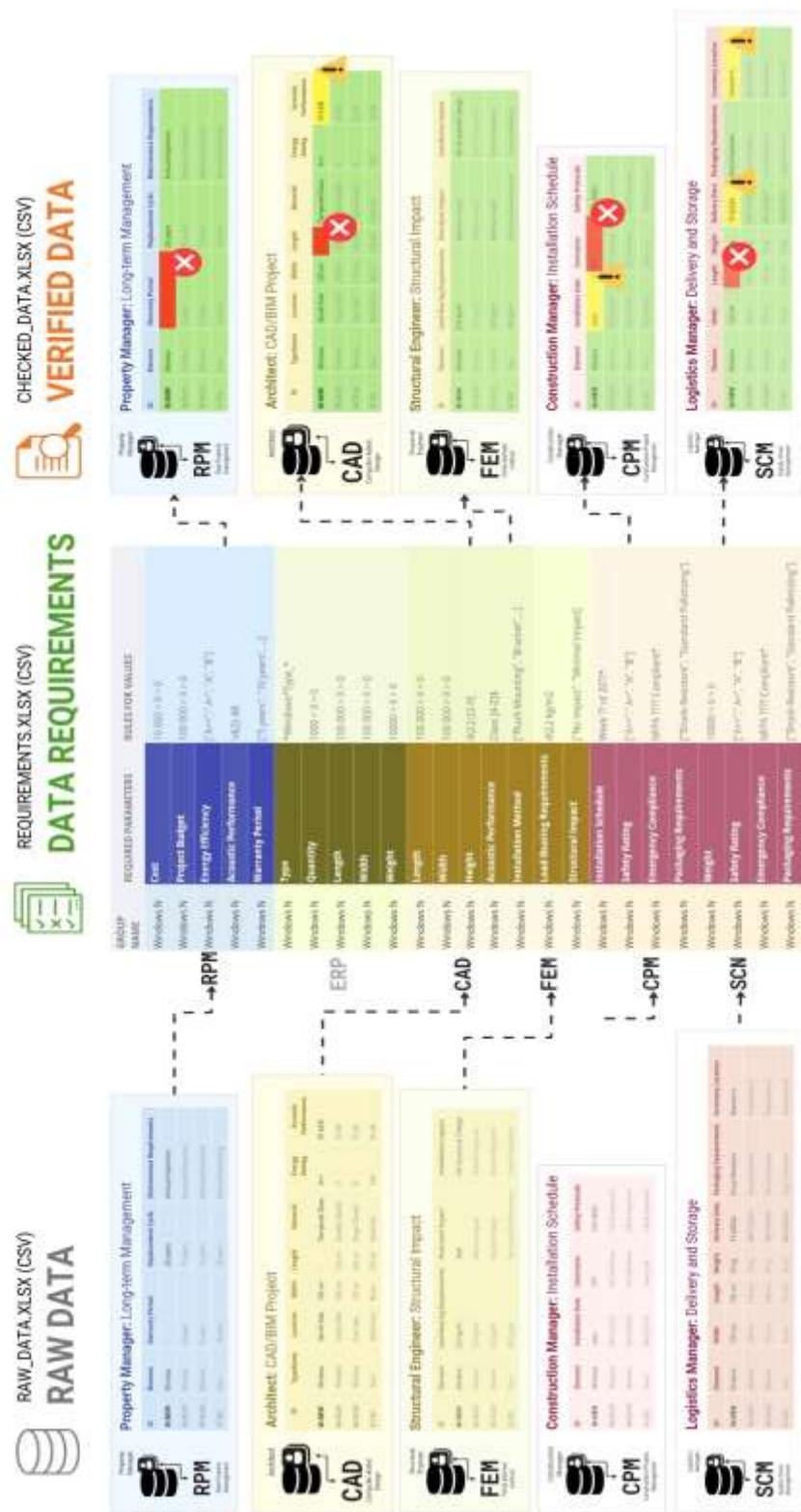
ID	Type Name	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met	
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B-	30	True

4.4-15. ábra Ellenőrzés azonosítja azokat az entitásokat, amelyek nem mentek át az ellenőrzési folyamatban, és egy új attribútumot ad hozzá az eredményekhez "Hamis" vagy "Igaz" értékekkel.

Hasonlóképpen, a projekthez való ablak hozzáadásának folyamata során a különböző szakemberektől kapott összes adatban (4.4-1. ábra) ellenőrizzük az összes projektelem (entitás) és az egyes rendszerek, táblázatok vagy adatbázisok szükséges attribútumainak konziszenciáját (4.4-1. ábra).

A végső táblázatban a szemléltetés érdekében célszerű az ellenőrzés eredményeit színnel kiemelni: zölddel jelöljük az ellenőrzésen sikeresen átment attribútumokat, sárgával a nem kritikus eltéréseket mutató értékeket, pirossal pedig a kritikus eltéréseket (4.4-16. ábra).

A validálás eredményeként (4.4-16. ábra) megkapjuk a megbízható és validált elemek listáját azonosítóikkal együtt, amelyek megfelelnek az attribútumkövetelményeknek. A validált elemek biztosítékot nyújtanak arra, hogy ezek az elemek megfelelnek a megadott szabványoknak és specifikációknak az Ablak osztály vagy bármely más osztály elemeinek hozzáadásában részt vevő valamennyi rendszer számára (az adatérvénysítés automatizálásáról és az automatizált ETL-folyamat létrehozásáról bővebben az "ETL és adatérvénysítés automatizálása" című fejezetben tárgyalunk).



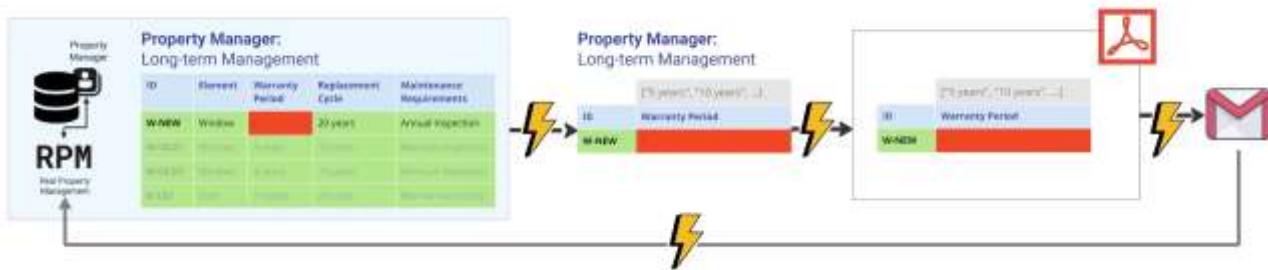
Ábra 4.4-16 Az összes rendszerre elvégzett ellenőrzés eredménye lehetővé teszi annak megállapítását, hogy mely adatok nem felelnek meg a vállalat követelményeinek.

A sikeresen validált entitások általában nem igényelnek sok figyelmet. Akadályok nélkül haladnak tovább a feldolgozás és más rendszerekbe való integrálás következő szakaszba. A "minőségi" elemekkel ellentétben a legnagyobb érdeklődésre azok az elemek tarthatnak számot, amelyek validálása nem sikerül. Az ilyen eltérésekről szóló információk kritikus fontosságúak: ezeket nemcsak táblázatos jelentésekben kell közölni, hanem különböző vizualizációs eszközökkel is. Az ellenőrzés eredményeinek grafikus ábrázolása segít az adatminőség általános állapotának gyors értékelésében, a problémás területek azonosításában és a korrekciós vagy javító intézkedések azonnali megtételében.

A ellenőrzési eredmények vizualizálása

A vizualizáció A a vizsgálati eredmények értelmezésének alapvető eszköze. A szokásos összefoglaló táblázatokon kívül tartalmazhat információs paneleket, diagramokat és automatikusan generált PDF dokumentumokat, amelyek a projektelemeket az ellenőrzési státuszuk szerint csoportosítják. A színkódolás itt támogató szerepet játszhat: a zöld jelezheti a sikeresen ellenőrzött elemeket, a sárga jelezheti a további figyelmet igénylő elemeket, a piros pedig a kritikus hibákat vagy a kulcsfontosságú adatok hiányát.

Példánkban (4.4-1. ábra) lépésről lépésre elemezzük az egyes rendszerek adatait: a CAD (BIM) és az ingatlankezeléstől a logisztikáig és a telepítési ütemtervekig (4.4-16. ábra). Az ellenőrzést követően az egyes szakemberek számára automatikusan egyedi riasztások vagy jelentésdokumentumok készülnek, például PDF formátumban (4.4-17. ábra). Ha az adatok helyesek, a szakember egy rövid üzenetet kap: "Köszönjük a közös munkát". Ha eltéréseket észlel, részletes jelentést küld a következő szöveggel: "Ez a dokumentum olyan elemeket, azok azonosítói, attribútumait és értékeit sorolja fel, amelyek megfelelőségét nem ellenőriztük".



4.4-17. ábra A validálás és az automatikus jelentéskészítés felgyorsítja az adathiányok feltárásának és megértésének folyamatát az adatokat létrehozó szakember számára.

Az automatikus érvényesítési folyamatnak köszönhetően - amint hibát vagy adathiányt észlel - azonnali értesítést küldünk chat-üzenet, e-mail vagy PDF -dokumentum formájában az érintett entitások és attribútumaik létrehozásáért vagy feldolgozásáért felelős személynek (4.4-18. ábra), a nem érvényesített elemek és attribútumleírások listájával együtt.



Ábra 4.4-18 Az automatikus ellenőrzési jelentések megkönnyítik a hibák megértését és felgyorsítják a projektadatok kitöltésével kapcsolatos munkát.

Ha például egy dokumentum (strukturálás után) úgy érkezik az ingatlankezelő rendszerbe, hogy a "Garanciaidőszak" attribútum hibásan van kitöltve, az ingatlankezelő riasztást kap az ellenőrizendő és javítandó attribútumok listájával.

Hasonlóképpen, a telepítési ütemterv vagy a logisztikai adatok bármilyen hiányossága automatikus jelentést generál, és például egy chat- vagy e-mail-értesítést küld az ellenőrzés eredményéről az illetékes szakembernek.

Az eredményeket tartalmazó PDF -dokumentumok és grafikonok mellett lehetőség van műszerfalak és interaktív 3D -modellek (7.1-6. ábra, 7.2-12. ábra) létrehozására, amelyek kiemelik a hiányzó attribútumokkal rendelkező elemeket, lehetővé téve a felhasználók számára, hogy vizuálisan használják az elemek 3D geometriáját a projektben szereplő elemek minőségének és teljességének szűrésére és értékelésére.

A vizsgálati eredmények vizualizálása automatikusan generált dokumentumok, grafikonok vagy műszerfalak formájában nagymértékben leegyszerűsíti az adatok értelmezését és megkönnyíti a projekt résztvevői közötti hatékony kommunikációt.

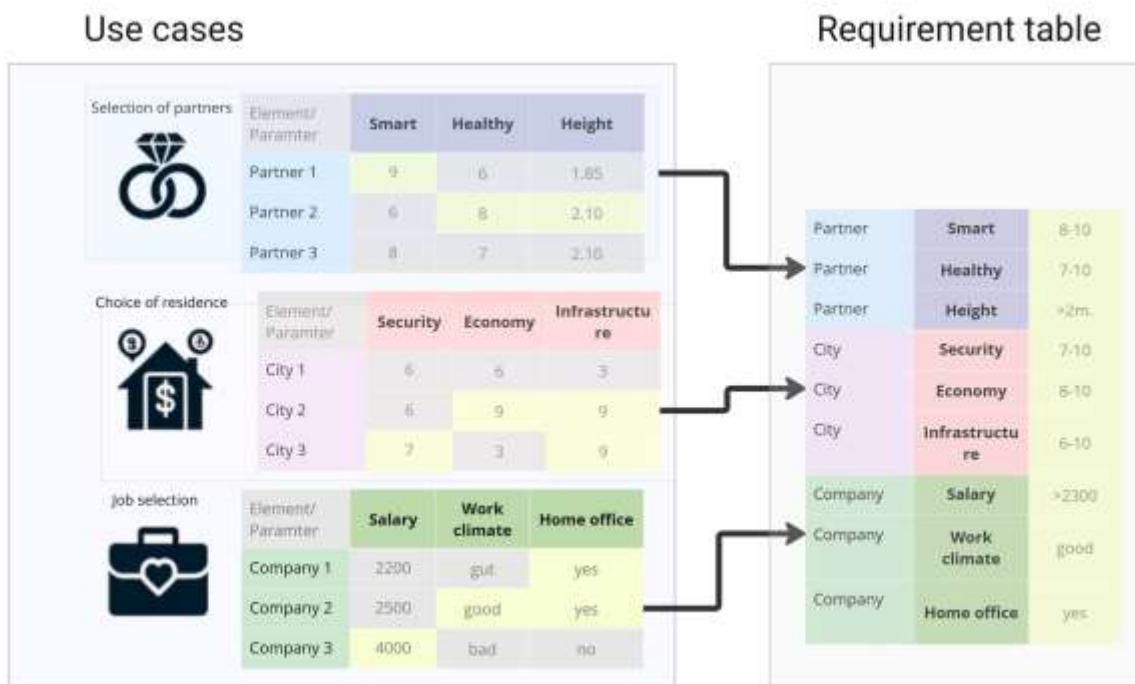
A különböző rendszerekből és információforrásokból származó adatok automatikus ellenőrzésének folyamata a minden nap életben alkalmazott megalapozott döntéshozatalhoz hasonlítható. Ahogy az építőipari vállalatok számos változót mérlegelnek - a bemeneti adatok megbízhatóságától kezdve a projekt megvalósításának időzítésére, költségeire és minőségére gyakorolt hatásukig -, ugyanígy az ember fontos döntések meghozatalakor, például lakóhelyválasztáskor számos tényezőt mérlegel: közlekedési elérhetőség, infrastruktúra, költségek, biztonság, életminőség. Mindezek a szempontok olyan szempontrendszer alkotnak, amely az életünket meghatározó végső döntések alapját képezi.

Az adatminőség-ellenőrzések összehasonlítása az emberi élet szükségleteivel

Az adatminőség-ellenőrzési módszerek és eszközök folyamatos fejlődése ellenére az információ megfelelőségének alapelve változatlan marad. Ez az elv egy kiforrott irányítási rendszer alapjaiba épül be, akár az üzleti életben, akár a minden nap életben.

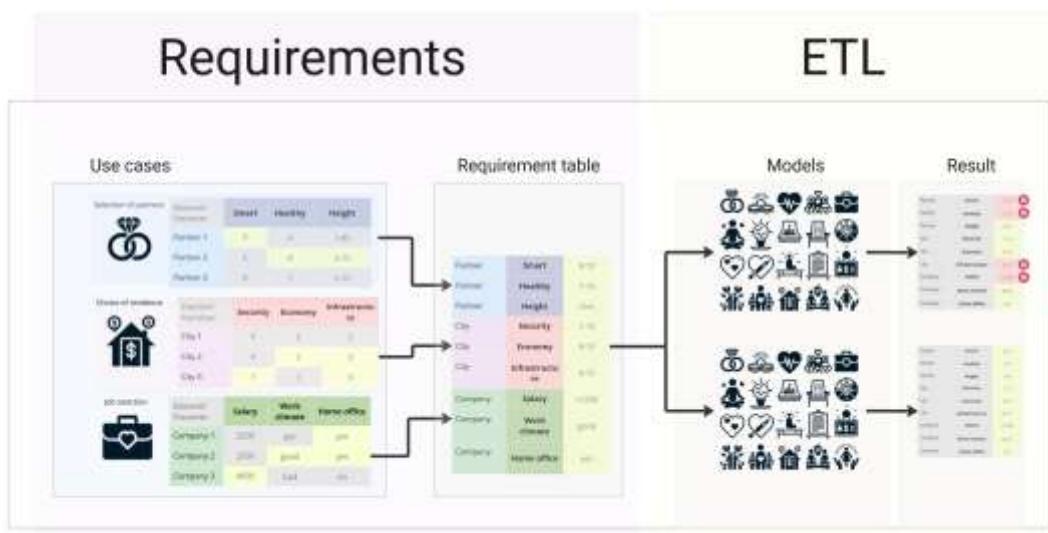
Az iteratív adatérvényesítés folyamata nagyban hasonlít ahhoz a döntéshozatali folyamathoz, amellyel mindenki nap mint nap szembesül. Mindkét esetben a tapasztalatra, az adatokra és a rendelkezésre álló új információra támaszkodunk. És egyre több életbeli és szakmai döntés - a stratégiai döntésekkel a minden nap döntésekig - születik adatok alapján.

Amikor például lakóhelyet vagy élettársat választunk, intuitív módon kialakítunk magunkban egy táblázatot a kritériumokról és jellemzőkről, amelyek alapján összehasonlítjuk az alternatívákat (4.4-19. ábra). Ezek a jellemzők - legyenek azok egy személy személyes tulajdonságai vagy egy ingatlan paraméterei - olyan attribútumokat képviselnek, amelyek befolyásolják a végső döntést.



Ábra: 4.4-19 A lakóhely, a munkaviszony vagy az élettársi kapcsolat megválasztása az egyéni attribútumkövetelményeken alapul.

A strukturált adatok használata és a követelmények leírásának formalizált megközelítése (4.4-20. ábra) hozzájárul a megalapozottabb és tájékozottabb döntésekhez mind a szakmai, mind a magánéletben.

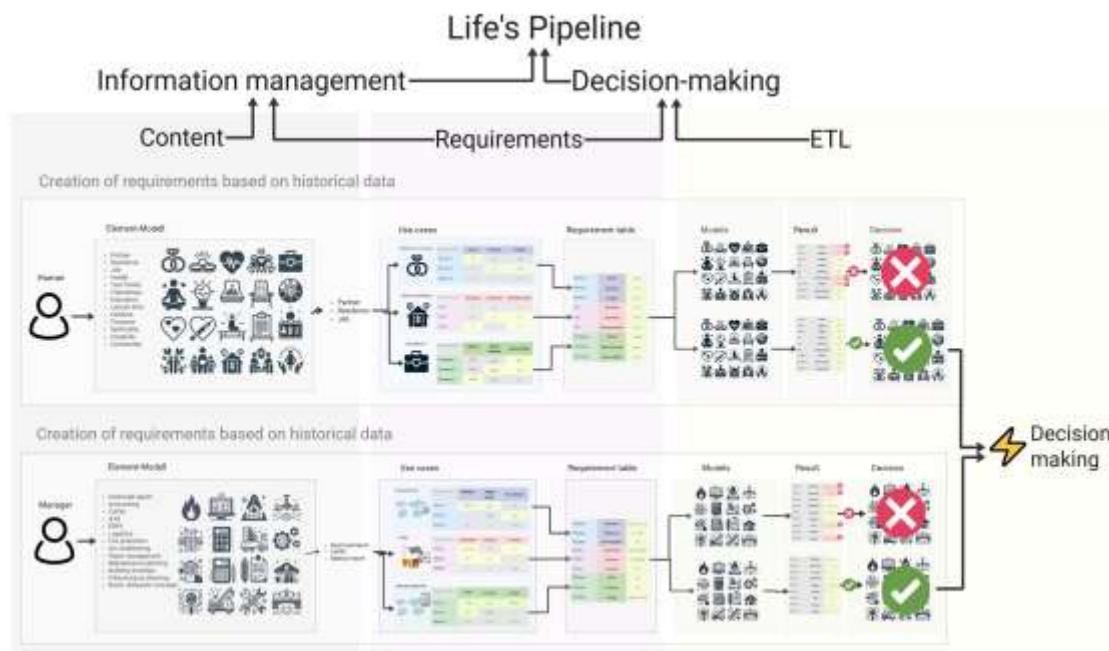


Ábra: 4.4-20 A követelmények formalizálása segít rendszerezni az életfelfogást és az üzleti döntéseket.

Az adatvezérelt döntéshozatali megközelítés nem kizárálag üzleti eszköz. A minden nap életbe is zökkenőmentesen beépül, az ETL-folyamathoz (Extract, Transform, Load) hasonlóan, a közös adatfeldolgozási lépéseket követve (4.4-21. ábra), amelyeket az adatok strukturálásánál már tárgyaltunk e rész elején, és amelyeket a könyv hetedik részében a feladatautomatizálás kontextusában részletesen megvizsgálunk:

- **Az adat mint alap (kivonat):** Bármely területen - legyen az a munka vagy a magánélet - információkat gyűjtünk. Az üzleti életben ezek lehetnek jelentések, számok, piaci adatok; a magánéletben személyes tapasztalatok, szeretteink tanácsai, visszajelzések, megfigyelések.
- **Értékelési kritériumok (Transform):** az összegyűjtött információkat előre meghatározott kritériumok alapján értelmezik. A munkahelyen ezek a teljesítménymutatók (KPI-k), költségvetési korlátok és normák; a magánéletben olyan paraméterek, mint az ár, a kényelem, a megbízhatóság, a karizma stb.
- **Előrejelzés és kockázatelemzés (Load):** az utolsó szakaszban a transzformált adatok elemzésén és a lehetséges következmények összehasonlításán alapuló döntéshozatalra kerül sor. Ez hasonló az üzleti folyamatokhoz, ahol az adatok üzleti logikán és kockázati szűrőn mennek keresztül.

A döntéseink - a triviális preferenciáktól kezdve, hogy mit együnk reggelire, egészen az olyan jelentős életeseményekig, mint a karrier vagy az élettárs kiválasztása - eredendően az adatok feldolgozásának és értékelésének eredményei.



4.4-21 Az üzleti élet és általában az élet adatvezérelt döntések sorozata, ahol a döntésekhez használt adatok minősége kulcsfontosságú tényező.

Életünkben minden mindenkel összefügg, és ahogyan az élő szervezetek, beleérte az embert is, a természet törvényeit követik, fejlődnek és alkalmazkodnak a változó körülményekhez, úgy az emberi folyamatok, beleérte az adatgyűjtést és -elemzést is, ezeket a természeti elveket tükrözik. A természet és az emberi tevékenység közötti szoros kapcsolat nemcsak a természettől való függőségünket erősíti meg, hanem azt is, hogy a több millió éves evolúció által csiszolt törvényeket alkalmazzuk a döntéshozatalhoz szükséges adatarchitektúrák, folyamatok és rendszerek létrehozására.

Az új technológiák, különösen az építőiparban, kiváló példája annak, hogy az emberiséget a természet újra és újra jobb, fenntarthatóbb és hatékonyabb megoldások létrehozására inspirálja.

Következő lépések: az adatok pontos számításokká és tervkké alakítása

Ebben a részben azt vizsgáltuk, hogyan lehet a strukturálatlan adatokat strukturált formátumba konvertálni, adatmodelleket kidolgozni és folyamatokat szervezni az építési projektek információminőségének ellenőrzésére. Az adatkezelés, szabványosítás és osztályozás alapvető folyamat, amely szisztematikus megközelítést és az üzleti követelmények világos megértését igényli. Az ebben a részben tárgyalt technikák és eszközök lehetővé teszik a különböző rendszerek közötti megbízható integrációt az objektum teljes életciklusa során.

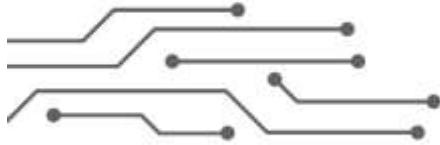
Összefoglalva ezt a részt, emeljük ki a főbb gyakorlati lépéseket, amelyek segítenek a tárgyalt megközelítések alkalmazásában a napi feladatok során:

- Kezdje a követelmények rendszerezésével
- Létrehozhatja a projektek és a folyamatok kulcselemeihez tartozó attribútumok és paraméterek nyilvántartását.

- Dokumentálja az egyes attribútumok határértékeit
- Folyamatok és az osztályok, rendszerek és attribútumok közötti kapcsolatok vizualizálása folyamatábrák segítségével (pl. Miro, Canva, Visio).
- Adatkonverzió automatizálása**
 - Ellenőrizze, hogy a folyamatokban gyakran használt dokumentumai közül melyek digitalizálhatók OCR-könyvtárak segítségével és alakíthatók át táblázatos formába.
 - Tekintse meg a fordított mérnöki eszközököt a CAD (BIM) adatainak kinyeréséhez.
 - Próbálja meg beállítani az automatikus adatlekérdezést a munkája során gyakran használt dokumentumokból vagy formátumokból egy táblázatos formába.
 - Automatikus konverziók beállítása különböző adatformátumok között
- Tudásbázis létrehozása a kategorizáláshoz**
 - Belső vagy meglévő elemosztályozó kifejlesztése vagy használata az ipari szabványokhoz igazodóan.
 - Dokumentálja a különböző osztályozási rendszerek közötti összefüggéseket.
 - Beszélje meg csapatával az elemek egységes azonosítási rendszerének és egyértelmű osztályozásának témáját.
 - Kezdje el egy olyan folyamat kialakítását, amely automatikusan validálja az adatokat - minden azokat, amelyekkel a csapaton belül dolgozik, minden azokat, amelyeket külső rendszereknek továbbítanak.

E megközelítések alkalmazásával jelentősen javíthatja az adatok minőségét, és egyszerűsítheti azok későbbi feldolgozását és átalakítását. A könyv következő részeiben azt vizsgáljuk meg, hogyan lehet a már strukturált és előkészített adatokat automatizált számításokhoz, költségbecsléshez, ütemezéshez és építési projektmenedzsmenthez alkalmazni.





V RÉSZ

KÖLTSÉG- ÉS IDŐSZÁMÍTÁS: AZ ADATOK BEÉPÍTÉSE AZ ÉPÍTÉSI FOLYAMATOKBA

Az ötödik rész az építési projektek költségszámításának és tervezésének optimalizálását szolgáló adatok felhasználásának gyakorlati szempontjaival foglalkozik. Részletesen elemzi a költségbecslés erőforrás-alapú módszerét és a becslési folyamatok automatizálását. Megvizsgálják a Quantity Take-Off (QTO) automatizált beszerzésének módszereit a CAD (BIM)-modellek ből és ezek integrációját a számítási rendszerekkel. A 4D és 5D modellezési technikákat az időtervezés és az építési költségek kezelése céljából vizsgálják, és konkrét példákat mutatnak be alkalmazásukra. A 6D -8D kiterjesztett információs rétegek elemzése, amelyek integrált megközelítést biztosítanak az ingatlanok fenntarthatóságának, működésének és biztonságának értékeléséhez. Az építési projektek szénlábjnyom- és ESG -indikátorainak számítási módszereit a modern környezetvédelmi követelmények és szabványok összefüggésében részletesen megvizsgálják. A hagyományos ERP és PMIS rendszerek és PMIS lehetőségei és korlátai az építési folyamatok irányításában kritikusan kerülnek értékelésre, elemezve az árképzés átláthatóságára gyakorolt hatásukat. A zárt megoldásokról a nyílt szabványokra és rugalmas adatelemző eszközökre való áttérés kilátásait jósolják, amelyek képesek az építési folyamatok nagyobb hatékonyságát biztosítani.

FEJEZET 5.1.

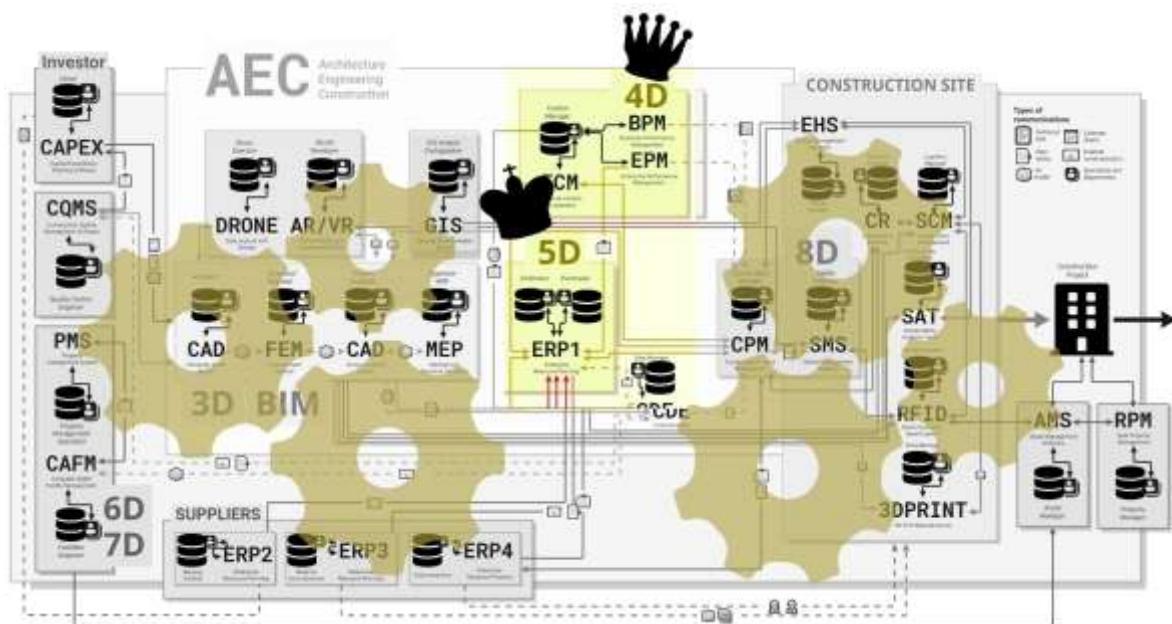
ÉPÍTÉSI PROJEKTEK KÖLTSÉGSZÁMÍTÁSAI ÉS BECSLÉSEI

Építőipari alapismeretek: mennyiségi-, költség- és időbecslés

Az építőipari vállalat fenntarthatóságát meghatározó számos üzleti folyamat közül - akárcsak évezredekkel ezelőtt - különösen fontosak az elemek számának, a projektköltségnek és a szállítási időnek a pontos becslésére szolgáló folyamatok (5.1-1. ábra).

Az írás kialakulása több tényező együttes hatására alakult ki, többek között azért, mert a korai társadalmakban szükség volt a gazdasági tranzakciók, a kereskedelelem és az erőforrás-gazdálkodás rögzítésére. Az első jogi jelentőségű dokumentumokat, az anyagköltségeket és munkadíjakat tartalmazó agyagtáblákat a kereskedeleml és az építkezés területén használták. Ezek a táblák rögzítették a felek kötelezettségeit az építkezések során, és a megállapodások, valamint a pénz- és kereskedelmi kapcsolatok bizonyítékként őrizték őket.

A becslés megközelítése évezredeken át nagyrészt változatlan maradt: a számításokat kézzel végezték, a becslőmérnök tapasztalatára és intuíciójára támaszkodva. A moduláris ERP-rendszerök és a CAD-eszközök megjelenésével azonban a mennyiségi-, költség- és időbecslés hagyományos megközelítése gyorsan átalakult. A mai digitális technológiák lehetővé teszik az idő és a költségek kulcsfontosságú számításainak teljes automatizálását, ami nagyobb pontosságot, gyorsaságot és átláthatóságot tesz lehetővé az építési projektek erőforrás-tervezésében.



5.1-1 ábra A sokféle rendszer közül a volumen-, költség- és időmérésért felelős eszközök a legfontosabbak az üzleti életben.

Az építőipari vállalatok a pontos idő- és költségadatokra összpontosítanak. Ezek viszont a felhasznált anyagok és munkaerő mennyiségétől függnek, átláthatóságuk pedig befolyásolja a nyereségességet. A számítási folyamatok bonyolultsága és az átláthatóság hiánya azonban gyakran vezet magasabb projektköltségekhez, határidő-kihagyáshoz, sőt csődhöz.

A KPMG "Ismerős problémák - új megközelítések" című jelentése (2023) szerint az építési projekteknek csak 50%-a fejeződik be időben, és a vállalatok 87%-a számolt be a tőkeprojektek gazdaságosságának fokozott ellenőrzéséről. A fő problémák a szakképzett személyzet hiányával és a kockázat-előrejelzés nehézségével kapcsolatosak [2].

A múltbeli projektek építése során az építőipari vállalat egész életében gyűjtötték történeti költségszámítási és folyamatidő-adatokat, amelyeket különböző rendszerek adatbázisaiba (ERP, PMIS BPM, EPM stb.) vittek be.

A jó minőségű múltbeli költségszámítási adatok megléte jelentős versenyelőnyt jelent egy építőipari szervezet számára, amely közvetlenül befolyásolja túlélőképességét.

Az építőipari és mérnöki vállalatok becslési és költségszámítási részlegei azért jönnek létre, hogy összegyűjtsék, tárolják és frissítsék a projektszámításokra vonatkozó múltbeli adatokat. Fő feladatak a vállalat tapasztalatainak összegyűjtése és rendszerezése, ami lehetővé teszi, hogy az új projektek terjedelmének, ütemezésének és költségeinek becslési pontossága idővel javuljon. Ez a megközelítés segít minimalizálni a hibákat a már megvalósított projektek gyakorlatán és eredményein alapuló jövőbeli számítások során.

A projektek becsült költségeinek kiszámítási módszerei

A költségszámítási szakemberek számos becslési módszert alkalmaznak, amelyek mindegyike az adatok egy adott típusára, az információk elérhetőségére és a projekt részletezettségi szintjére összpontosít. A leggyakoribbak közé tartoznak:

- **Erőforrás-alapú módszer:** egy projekt becsült költségének becslése a szükséges erőforrások, például anyagok, berendezések és munkaerő részletes elemzése alapján. Ez a módszer megköveteli az összes feladat és az egyes feladatok elvégzéséhez szükséges erőforrások részletes felsorolását, majd ezek költségének kiszámítását. Ez a módszer rendkívül pontos, és széles körben használják a költségbecslésben.
- **Parametrikus módszer:** statisztikai modelleket használ a projekt paraméterei alapján történő költségbecsléshez. Ez magában foglalhatja a mértékegységenkénti költségek - például az épület területe vagy a munkaterület - elemzését, és ezen értékek hozzáigazítását a konkrét projektfeltételekhez. A módszer különösen hatékony a korai szakaszokban, amikor még nem állnak rendelkezésre részletes információk.
- **Egységmódszer (egysékgöltség-módszer):** a projekt becsült költségét egy mértékegységre (pl. négyzetméterre vagy köbméterre) vetített költség alapján számítja ki. Ez gyors és egyszerű módot biztosít a különböző projektek vagy projektrészek költségeinek összehasonlítására és elemzésére.
- **Szakértői megítélés (Delphi-módszer):** olyan szakértők véleményén alapul, akik a tapasztalataikat és ismereteiket felhasználva becsülik meg a projekt értékét. A megközelítés akkor hasznos, ha nem állnak rendelkezésre pontos kiindulási adatok, vagy a projekt egyedi.

Érdemes megjegyezni, hogy a parametrikus módszer és a szakértői ítéletek adaptálhatók a gépi tanulási modellekhez. Ez lehetővé teszi a projektköltség- és ütemterv-előrejelzések automatikus generálását a képzési minták alapján. Az ilyen modellek alkalmazására vonatkozó példákat részletesebben a "Példa a gépi tanulás alkalmazására projektköltségek és határidők megállapítására" című fejezetben tárgyaljuk (9.3-5. ábra).

Mindazonáltal az erőforrás-alapú módszer továbbra is a legnépszerűbb és a világ gyakorlatában legszélesebb körben alkalmazott módszer. Nemcsak a becsült költségek pontos értékelését biztosítja, hanem lehetővé teszi az építkezésen zajló egyes folyamatok időtartamának és a projekt egészének kiszámítását is (további részletek az "Építési ütemtervek és 4D -projektadatok" című fejezetben).

Erőforrás-alapú módszer költségszámítás és becslés az építőiparban

Az erőforrás-alapú költségszámítás a vezetői számvitel olyan módszere, amelyben egy projekt költségei az összes érintett erőforrás közvetlen elszámolásán alapulnak. Az építőiparban ez a megközelítés magában foglalja a munka elvégzéséhez szükséges összes anyagi, munka- és műszaki erőforrás részletes elemzését és értékelését.

Az erőforrás-alapú módszer, nagyfokú átláthatóságot és pontosságot biztosít a költségvetés-tervezésben, mivel az erőforrások tényleges áraira összpontosít a becslés időpontjában. Ez különösen fontos az instabil gazdasági környezetben, ahol az áringadozások jelentősen befolyásolhatják a projekt összköltségét.

A következő fejezetekben részletesen megvizsgáljuk az erőforrás-alapú költségszámítás folyamatát. Ahhoz, hogy jobban megértsük az építőiparban alkalmazott elveit, egy analógiát fogunk húzni egy étteremben elköltött vacsora költségszámításával. Az étterem vezetője az estét megtervezve összeállítja a szükséges termékek listáját, figyelembe veszi az egyes ételek főzési idejét, majd a költségeket megszorozza a vendégek számával. Az építőiparban a folyamat hasonló: a projektelemek (objektumok) minden egyes kategóriájára vonatkozóan tételes becslések készülnek Recipes, és a projekt teljes költségét úgy határozzák meg, hogy az összes költséget összegzik egy teljes számlában - a kategóriánkénti végső becslésben.

Az erőforrás-alapú megközelítés kulcsfontosságú és kezdeti szakasza a vállalat kezdeti adatbázisának létrehozása. A költségszámítás első szakaszában strukturált listát állítanak össze minden olyan tételről, anyagról, munkatípusról és erőforrásról, amellyel a vállalat az építési projektjei keretében rendelkezik - a raktárban lévő szögtől kezdve az emberek leírásáig, képzettségükön és órabérükön keresztül. Ezeket az információkat egy egységes "Építési erőforrások és anyagok adatbázisa" - egy táblázatos nyilvántartás, amely tartalmazza a nevekre, jellemzőkre, mértékegységekre és aktuális áráakra vonatkozó adatokat - rendszerezzi. Ez az adatbázis lesz a fő és elsődleges információforrás minden későbbi erőforrás-számításhoz - mind a munkák költségeihez, mind az ütemezéshez.

Építőipari erőforrások adatbázisa: építőanyagok és építési munkák katalógusa

Az építési erőforrások és anyagok adatbázisa vagy táblázata - részletes információkat tartalmaz minden egyes, az építési projektben felhasználható elemről - termékről, tételről, anyagról vagy szolgáltatásról,

beleértve annak nevét, leírását, mértékegységét és egységköltségét, strukturált formában rögzítve. Ebben a táblázatban a projektekben felhasznált különböző típusú tüzelőanyagoktól és anyagoktól kezdve a szakemberek részletes listájáig minden megtalálhat különböző kategóriák formájában, az óradíjak leírásával (5.1-2. ábra).

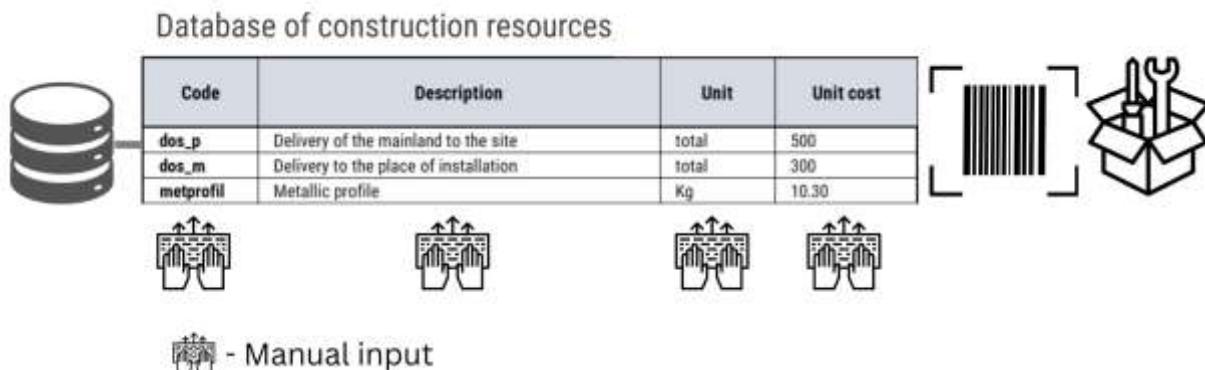
Database of resources	
 1st grade potatoes 1 kg \$2,99	 Sand lime bricks 1 pcs \$1
 Black Angus marble beef 1 kg \$26,99	 JCB 3CX backhoe loader 1 h \$150
 Broccoli 1 pcs \$1,99	 Laborer of the 1st category 1 h \$30

5.1-2 ábra Az erőforrástabla egy olyan összetevőlista, amely leírja az anyagokat és szolgáltatásokat egységköltséggel.

Az "erőforrás-adatbázis" egy webáruház termékkatalógusához hasonlít, ahol minden egyes termékhez részletes leírás tartozik, amely tartalmazza annak tulajdonságait. Ez megkönnyíti a költségbecslők számára a megfelelő erőforrások kiválasztását (mint a termékek kiválasztása a kosárba helyezéskor), amelyek a konkrét építési folyamatok számításaihoz szükségesek számítások formájában (végső megrendelés az online boltban).

Az erőforrás-adatbázis úgy is elképzelhető, mint egy éttermi szakácskönyv összes hozzávalójának listája. minden egyes építőanyag, berendezés és szolgáltatás hasonlít a receptekben használt összetevőkhöz. Az "erőforrás-adatbázis" az összes összetevő - építőanyag és szolgáltatás - részletes listája, beleértve az egységenkénti költségüket: darab, méter, óra, liter stb.

Az "Építőipari erőforrás-adatbázisok" táblába kétféleképpen lehet új entitáselemeket felvenni: manuálisan (5.1-3. ábra) vagy automatikusan, a vállalat készletgazdálkodási rendszereivel vagy beszállítói adatbázisokkal való integrálással.



Ábra 5.1-3 Az erőforrás-adatbázis kitöltése kézzel történik, vagy automatikusan átveszi az adatokat más adatbázisokból.

Egy tipikus középméretű építőipari vállalat olyan adatbázist használ, amely több ezer, néha több tízezer, az építési projektekben felhasználható, részletes leírással ellátott elemet tartalmaz. Ezeket az adatokat aztán automatikusan felhasználják a szerződésekben és a projektdokumentációban, hogy pontosan leírják a munka és a folyamatok összetételét

A változó piaci feltételek, például az infláció követése érdekében az erőforrás-adatbázisban (5.1-3. ábra) az egyes termékek (áruk vagy szolgáltatások) "egységköltség" attribútumát rendszeresen frissítik kézzel vagy más rendszerekből vagy online platformokról automatikusan letöltött aktuális árakkal.

Az erőforrás egységköltségének aktualizálása történhet havonta, negyedévente vagy évente - az erőforrás jellegétől, az inflációtól és a külső gazdasági környezettől függően. Az ilyen frissítésekre a számítások és becslések pontosságának fenntartása érdekében van szükség, mivel ezek az alapelemek képezik a költségbecslők munkájának kiindulópontját. A naprakész adatok segítségével olyan becslések, költségvetések és ütemtervek készülnek, amelyek tükrözik a valós piaci viszonyokat, és csökkentik a későbbi projektszámítások hibakockázatát.

Számítások összeállítása és a munkák költségszámítása az erőforrásbázis alapján.

Miután feltöltötte az "Építési erőforrás-adatbázist" (5.1-3. ábra) minimális egységegyésekkel, elkezdheti a számításokat létrehozni, amelyek az építkezésen végzett minden egyes folyamatra vagy munkára vonatkozóan bizonyos mértékegyiségekre vonatkozóan kerülnek kiszámításra: például egy köbméter betonra, egy négyzetméter gipszkartonfalra, egy méter szegélyre vagy egy ablak beépítésére.

Például egy 1 m²-es téglafal építéséhez (5.1-4. ábra), korábbi projektek tapasztalatai alapján, körülbelül 65 téglá (egység "Szilikát téglá") szükséges, darabonként 1 dollárért (attribútum "Darabonkénti költség"), összesen 65 dollárért. Továbbá, tapasztalataim szerint 10 percig építőipari berendezés ("JCB 3CX Loader" entitás) használatára van szükség, amely a téglákat a munkaterület közelében helyezi el. Mivel a berendezés bérlete 150 dollárba kerül óránként, a 6 perces használat körülbelül 15 dollárba kerülne. Ezen kívül 2 órára szükség lesz egy téglafektető vállalkozóra, akinek az órádja 30 dollár, összesen 60 dollár.

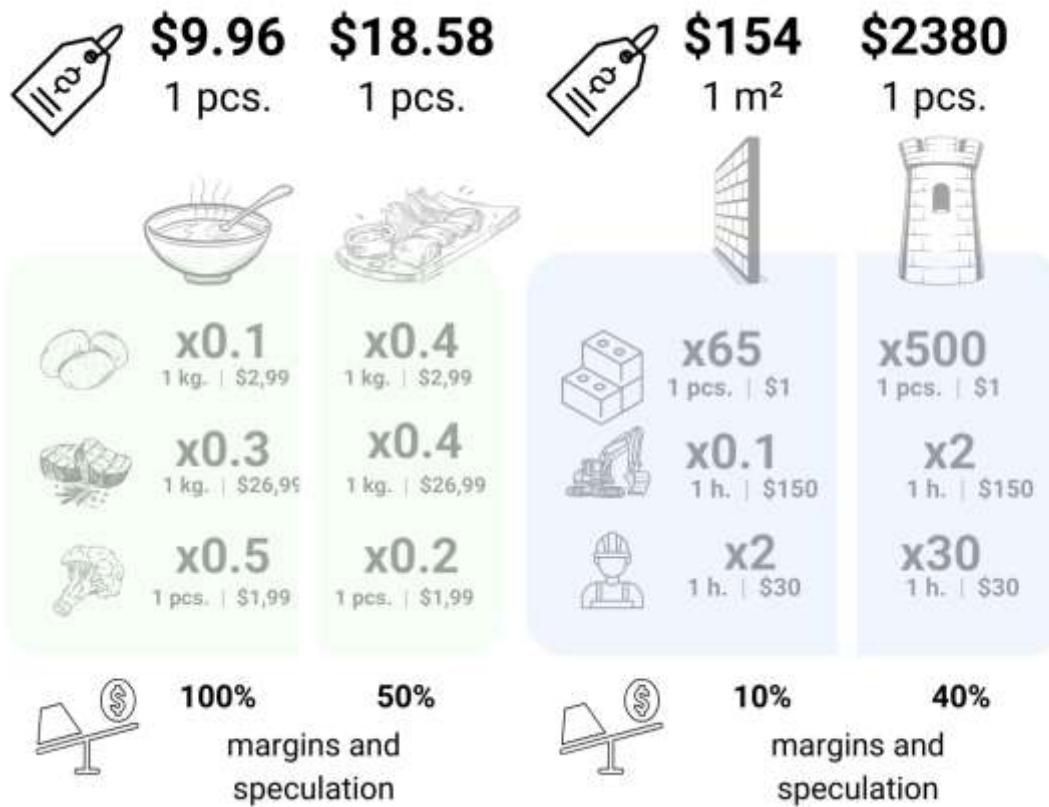
x0.1 1 kg \$2,99	x0.4 1 kg. \$2,99	x65 1 pcs \$1	x500 1 pcs \$1
x0.3 1 kg \$26,99	x0.4 1 kg. \$26,99	x0.1 1 h \$150	x2 1 h \$150
x0.5 1 pcs \$1,99	x0.2 1 pcs. \$1,99	x2 1 h \$30	x30 1 h \$30
+ margins and speculation	+ margins and speculation	+ margins and speculation	+ margins and speculation

Ábra: 5.1-4 A költségszámítások részletes listát adnak a munkák és folyamatok elvégzéséhez szükséges építőanyagokról és szolgáltatásokról.

A számítások összetétele (az úgynevezett "receptek") a vállalat által a nagy mennyiségi hasonló munka elvégzése során felhalmozott történelmi tapasztalatok alapján alakul ki. Ezt a gyakorlati tapasztalatot általában az építkezésről érkező visszajelzések révén halmozzák fel. Különösen a művezető közvetlenül az építkezésen gyűjt az információkat, rögzítve a tényleges munkaerőköltségeket, az anyagfelhasználást és a technológiai műveletek árnyalatait. A becslési osztályal együttműködve ezeket az információkat aztán iteratív módon finomítják: a folyamatleírásokat finomítják, az erőforrás-összetételt kiigazítják, és a költségszámításokat frissítik a közelmúltbeli projektek tényleges adatainak figyelembevételevel.

Ahogyan egy recept leírja az étel elkészítéséhez szükséges összetevőket és mennyiségeket, úgy egy becslőlap is részletes listát ad egy adott munka vagy folyamat elvégzéséhez szükséges építőanyagokról, erőforrásokról és szolgáltatásokról.

A rendszeresen végzett munka lehetővé teszi a munkások, a művezetők és a becslők számára, hogy tájékozódjanak a szükséges erőforrás-mennyiségen: anyagok, üzemanyag, munkaidő és egyéb paraméterek, amelyek egy munkaegység elvégzéséhez szükségesek (5.1-5. ábra). Ezeket az adatokat táblázatok formájában viszik be a becslőrendszerükbe, ahol minden feladatot és műveletet az erőforrásbázis minimális elemein keresztül írnak le (folyamatosan frissített árakkal), ami biztosítja a számítások pontosságát.



Ábra 5.1-5 Az egységárakat minden egyes munkához összegyűjtik, ahol az egység mennyiségi attribútumát megszorozzák a mennyiséggel, és hozzáadnak egy profitszázeléket.

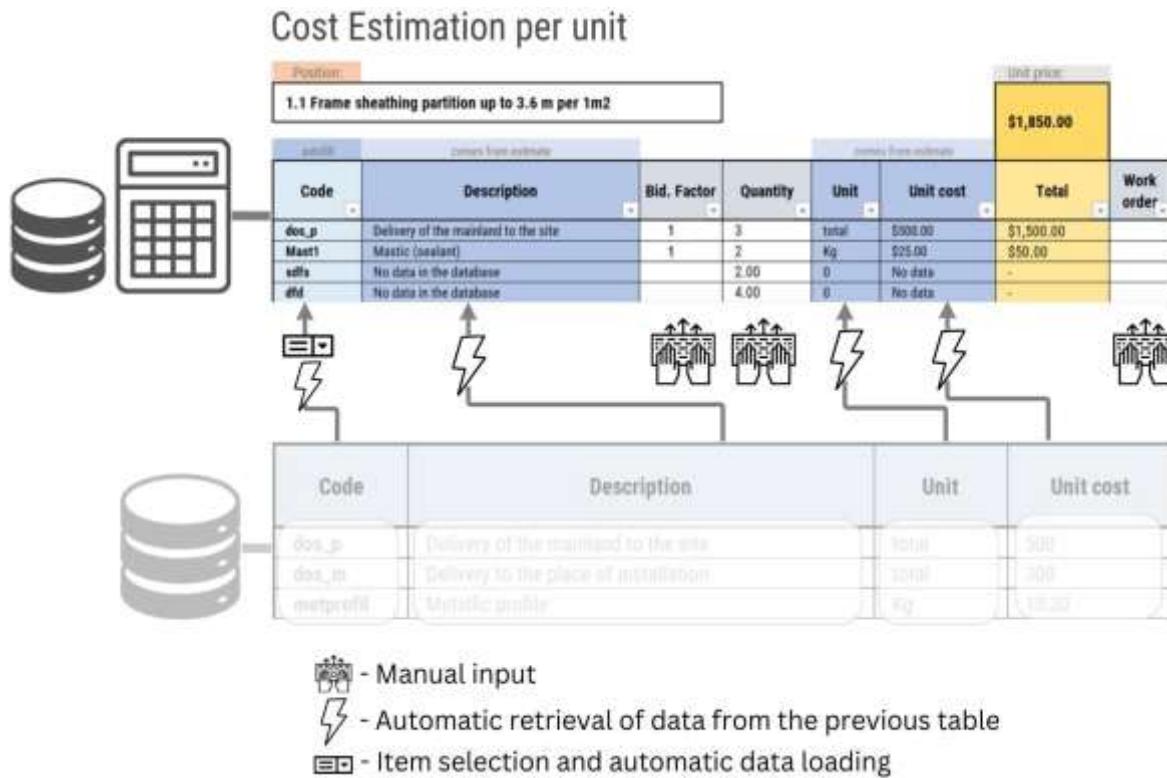
Az egyes folyamatok vagy tevékenységek (kalkulációs objektumok) összköltségének kiszámításához a költségattribútumot meg kell szorozni annak számával és tényezővel. Az együtthatók különböző tényezőket vehetnek figyelembe, például a munka összetettségét, a regionális jellemzőket, az inflációs rátát, a lehetséges kockázatokat (várható általános költséghányad) vagy a spekulációt (kiegészítő nyereségtényező).

A becslő, mint elemző, a művezető tapasztalatait és ajánlásait szabványosított becslésekkel alakítja át, táblázatos formában leírva az építési folyamatokat az erőforrás-egységeken keresztül. A becslő feladata lényegében az, hogy paraméterek és együtthatók segítségével összegyűjtse és strukturálja az építkezésről érkező információkat.

Így a munkaegységre (pl. négyzetmétere vagy köbmétere, vagy egy egység egy installációjára) jutó végső költség nemcsak az anyagok és a munkaerő közvetlen költségeit tartalmazza, hanem a vállalati felárakat, általános költségeket, biztosítási és egyéb tényezőket is (5.1-6. ábra)

Ugyanakkor nem kell többé aggódnunk a tényleges árak miatt a (recept) számítások során, mivel a valós árak minden tükröződnek az "erőforrásbázisban" (összetevő táblázat). A számítások szintjén az erőforrásbázisból automatikusan betöltődnek a táblázatba az adatok (pl. az elemkód vagy annak egyedi azonosítója alapján), amelyek betöltik a leírást és a tényleges egységárat, amely viszont automatikusan

betölthető az online platformokról vagy az építőanyagok online boltjából. A munkák számításainak szintjén a becslőnek csak a munkát vagy folyamatot kell leírnia az "erőforrások mennyisége" attribútumon és további tényezőkön keresztül.



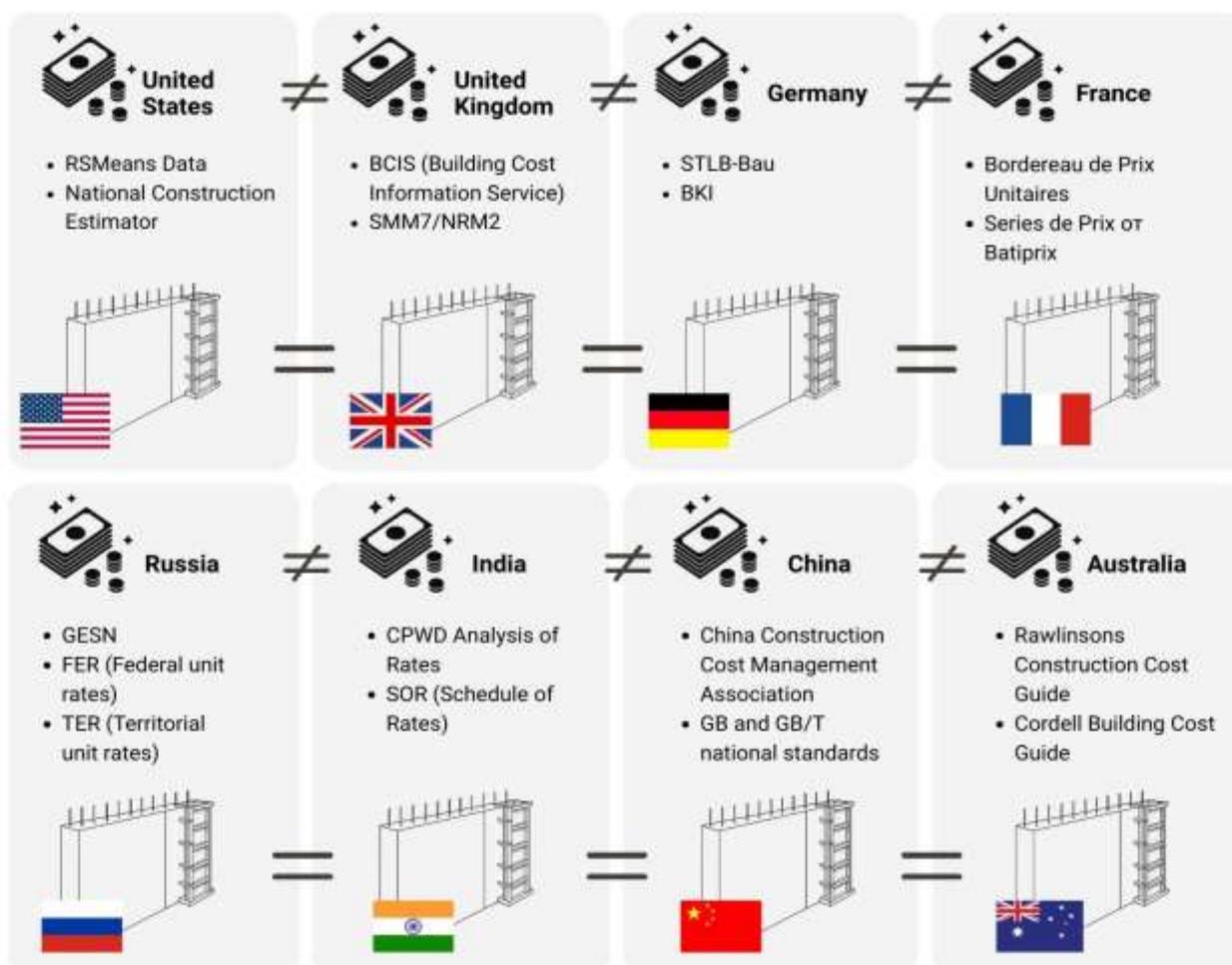
Ábra 5.1-6 A munkaegységköltség kiszámításának szakaszában csak a szükséges erőforrások számának attribútumait kell kitölteni, minden más automatikusan betöltődik az erőforrás-adatbázisból.

A létrehozott munkaköltségek tipikus projektek sablon táblázata formájában kerülnek tárolásra, amelyek közvetlenül kapcsolódnak az építési erőforrások és anyagok adatbázisához. Ezek a sablonok a jövőbeli projektek ismétlődő munkatípusainak szabványosított receptjeit jelentik, biztosítva a számítások egységességét a vállalaton belül.

Ha bármely erőforrás költsége változik az adatbázisban (5.1-3. ábra) - akár manuálisan, akár automatikusan, az aktuális piaci árak letöltésével (pl. inflációs körülmények között) -, a frissítések azonnal megjelennek az összes kapcsolódó költségszámításban (5.1-6. ábra). Ez azt jelenti, hogy csak az erőforrás-alapot kell módosítani, míg a költségszámítási sablonok és becslések idővel változatlanok maradnak. Ez a megközelítés biztosítja a számítások stabilitását és reprodukálhatóságát az esetleges áringadozások esetén, amelyeket csak egy viszonylag egyszerű erőforrástáblázatban vesznek figyelembe (5.1-3. ábra).

Minden egyes új projekthez a standard költségszámítási sablon egy másolata készül, ami lehetővé teszi, hogy a vállalat által elfogadott eredeti sablon megváltoztatása nélkül változtatásokat hajtson végre és a tevékenységeket az egyedi követelményekhez igazítsa. Ez a megközelítés rugalmasságot biztosít a számítások kiigazításában: figyelembe veheti az építkezés sajátosságait, az ügyfél kívánságait, bevezetheti a kockázati vagy jövedelmezőségi (spekulációs) együtthatókat - mindezt anélkül, hogy megsértené a vállalat szabványait. Ez segít a vállalatnak megtalálni az egyensúlyt a profitmaximalizálás, az ügyfelek elégedettsége és a versenyképesség fenntartása között.

Egyes országokban az ilyen, évtizedek alatt felhalmozott költségszámítási sablonokat nemzeti szinten szabványosítják, és az építési költségszámítási rendszer nemzeti szabványainak részévé válnak (5.1-7. ábra).



Ábra 5.1-7 A világ különböző országai saját költségszámítási szabályokkal rendelkeznek, saját (előíró) kompendiumokkal és szabványokkal az építőipari munkákra vonatkozóan, ugyanazon elem költségszámítására.

Az ilyen szabványosított becsült erőforrás-alapokat (5.1-7. ábra) minden piaci kötelező használni, különösen az államilag finanszírozott projektek esetében. Az ilyen szabványosítás biztosítja az átláthatóságot, az összehasonlíthatóságot és a méltányosságot az árak és a szerződéses kötelezettségek kialakításában az ügyfél számára

A projekt végső költségszámítása: a becsléstől a költségvetésig

Az állami és ágazatspecifikus becslési szabványok különböző szerepet játszanak a különböző országok építési gyakorlatában. Míg egyes országok megkövetelik egy egységes szabvány szigorú betartását, a legtöbb fejlett gazdaság rugalmassabb megközelítést alkalmaz. A piacgazdaságokban az állami építési szabványok általában csak alapként szolgálnak. Az építőipari vállalatok ezeket a szabványokat saját működési modelljükhez igazítják, vagy teljesen átdolgozzák, kiegészítve azokat saját, testre szabott tényezőkkel. Ezek a kiigazítások tükrözik a vállalati tapasztalatokat, az erőforrás-gazdálkodás hatékonyságát és gyakran olyan tényezőket, amelyekben például a vállalat spekulatív nyereségét is figyelembe veszik.

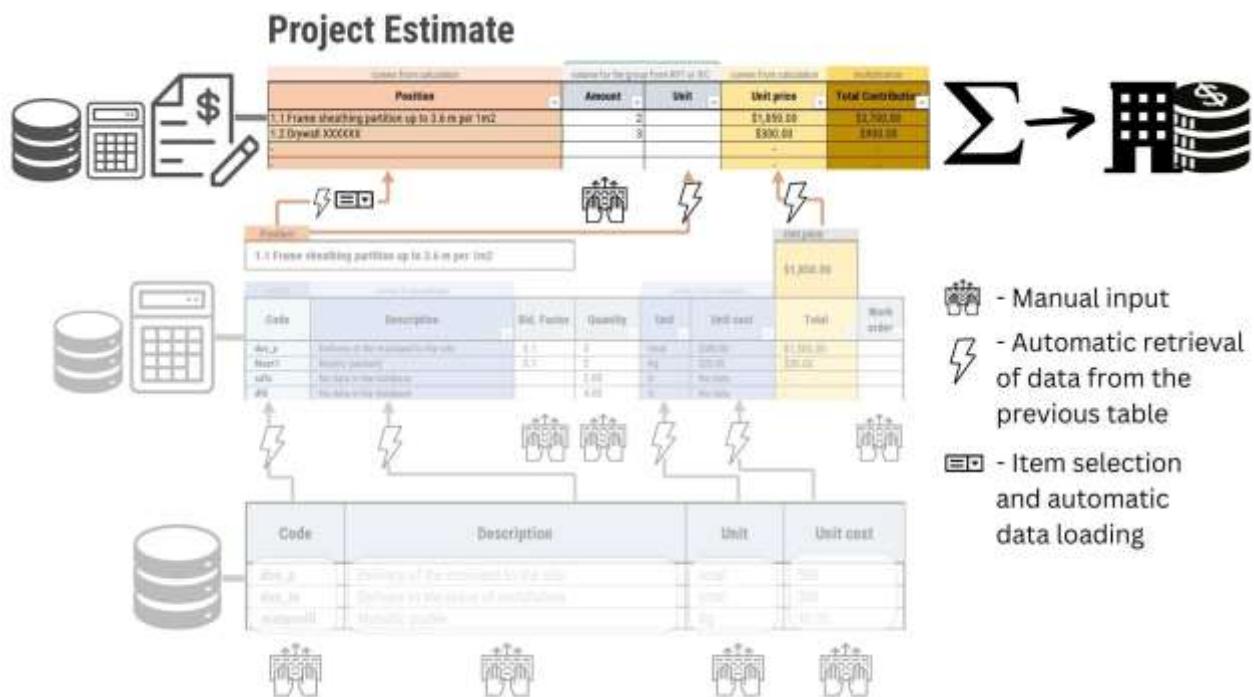
Ennek eredményeképpen a verseny szintje, a piaci kereslet, a megcélzott árrések és még az egyes ügyfelekkel való kapcsolatok is jelentős eltérésekhez vezethetnek a szabványosított normáktól. Ez a gyakorlat piaci rugalmasságot biztosít, ugyanakkor megnehezíti a különböző vállalkozók ajánlatainak átlátható összehasonlítását, és a kalkulációs folyamat e szakaszában a spekulatív árképzés elemét vezeti be az építőiparba.

Miután elkészültek az egyes tevékenységek és folyamatok számítási sablonjai - vagy gyakrabban egyszerűen lemásolták a szabványos kormányzati becslésekből (5.1-7. ábra), az adott vállalat "sajátosságait" tükröző együtthatókkal kiegészítve -, az utolsó lépés az egyes tételek költségének megszorzása az új projekt munkaterületének vagy folyamatainak megfelelő attribútumával.

Egy új építési projekt összköltségének kiszámításakor a legfontosabb lépés az összes költségszámítási téTEL költségeinek összegzése, megszorozva a projektben e munkárészek mennyiségével.

A projekt összköltségének létrehozásához egyszerűsített példánkban egy négyzetméter fal építési költségének kiszámításával kezdünk, és a számítás költségét (pl. "1m² falelemek szabványos beépítése" munka) megszorozzuk a projektben lévő összes négyzetméter falszámmal (pl. a "Falelemek" típusú entitás "Terület" vagy "Mennyiség" attribútuma (5.1-8. ábra) a projekt CAD oldaláról vagy a művezető számításaiból).

Hasonlóképpen kiszámítjuk a projekt összes elemének költségét (5.1-8. ábra): vesszük az egységnyi munkára jutó költséget, és megszorozzuk a projekt adott elemének vagy elemcsoportjának mennyiségével. A becslőnek csak a projektben lévő ilyen elemek, tevékenységek vagy folyamatok számát kell megadni a oldalon volumen vagy mennyiség formájában. Ez lehetővé teszi a teljes építési becslés automatikus generálását.



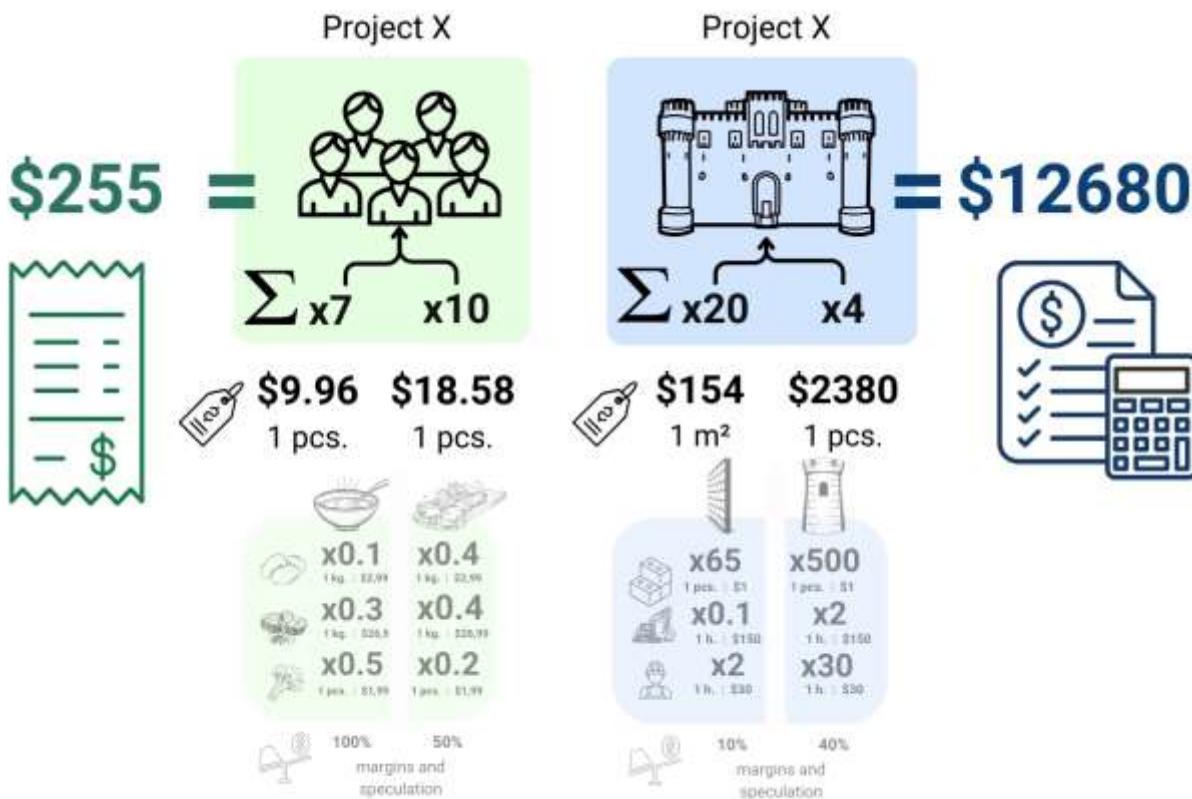
Ábra 5.1-8 A becslés készítésének szakaszában csak a munkaterületet adjuk meg.

A kalkulációkhöz hasonlóan ezen a szinten is automatikusan feltöljük a kész kalkulált tételeket (a kalkulációs sablonból vagy a sablonból másolt és szerkesztett új tételeket), amelyek automatikusan magukkal hozzák az aktuális munkaegységenkénti költséget (amely automatikusan frissül az erőforrás-adatbázisból (5.1-8. ábra alsó táblázat)). Ennek megfelelően az erőforrás-adatbázisban vagy a kalkulációs táblákban lévő adatok bármilyen változása esetén - a becslés adatai automatikusan frissülnek az aktuális napra vonatkozóan, anélkül, hogy magát a kalkulációt vagy a becslést módosítani kellene.

Éttermi kontextusban az esemény végső költségét hasonló módon számítják ki, és az megegyezik a teljes vacsora végső költségével, ahol az egyes ételek költsége és a vendégek száma szorozva a csekk összköltségét adja ki (5.1-9. ábra). És akár csak az építőiparban, az éttermi főzés receptjei sem változnak évtizedekig. Nem úgy, mint az árak, ahol az összetevők ára óránként változhat.

Ahogyan egy étteremtulajdonos megszorozza az egyes ételek költségét az adagok és a személyek számával, hogy meghatározza az esemény teljes költségét, a költségbecslésért felelős vezető összeadja az összes projektelem költségét, hogy teljes építési becslést kapjon.

Így a projekt minden egyes tevékenységére meghatározzuk annak végső költségét (5.1-9. ábra), amely megszorozva az adott tevékenységhez tartozó egység attribútumvolumenével - megadja a tevékenységcsoportok költségét, amelyből az egész projekt végső költségét kapjuk.



5.1-9 ábra A végső becslést úgy számoljuk ki, hogy az egyes elemek munkaköltség-attribútumait összeadjuk a terjedelem-attribútummal.

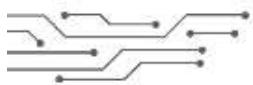
A projekt teljes költsége (5.1-8. ábra) pénzügyi képet ad a projektről, lehetővé téve az ügyfelek, befektetők vagy finanszírozó szervezetek számára, hogy megértsék a projekt megvalósításához szükséges teljes költségvetést és pénzügyi forrásokat egy adott napon, az aktuális árok figyelembevételével.

És ha az erőforrásbázisok, számítások és becslések (folyamatreceptek) összeállításának folyamatait már több tízezer éve kidolgozták, félautomatizálták és csiszolták, és állami szinten rögzítették, akkor a végső becslés utolsó szakaszához szükséges elemek mennyiségrére és mennyiségrére vonatkozó minőségi információk automatikus megszerzése - ma még minden szűk keresztmetszet a projekt költség- és időattribútumainak, és általában a projekt teljes költségvetésének összes számítási folyamatában.

A térfogatszámítás hagyományos módszere évezredek óta a térfogat és a mennyiség kézzel, sík rajzok segítségével történő mérése. A digitális korszak beköszöntével a vállalatok felfedezték, hogy a térfogat- és mennyiségi információk ma már automatikusan kinyerhetők a CAD-modellekben szereplő geometriai adatokból, ami forradalmasítja a mennyiségi adatok megszerzésének évezredes módszereit.

A folyamatbecslés és a becslés modern megközelítései magukban foglalják a térfogati és mennyiségi attribútumok automatikus kinyerését a CAD-adatbázisokból, amelyek feltölthetők és a oldalon összekapcsolhatók a költségszámítási folyamattal, hogy a tervezéstől az üzemeltetésig bármelyik

szakaszban naprakész projektcsoport-volumeneket kapjanak.



FEJEZET 5.2.

MENNYISÉGI SZÁMBAVÉTEL ÉS AUTOMATIKUS BECSLÉSEK ÉS ÜTEMTERVEK KÉSZÍTÉSE

A 3D-ről a 4D-re és 5D-re való áttérés: volumetrikus és mennyiségi paraméterek használata

A költségkalkulációs táblázatok és a leírt folyamatok erőforráson keresztül (5.1-8. ábra) birtokában a következő lépés a számításokhoz és a végső becsléshez szükséges elemek egy csoportjára vonatkozó mennyiségi vagy mennyiségi paraméterek automatikus megszerzése.

A projektelemek - pl. falak vagy födémek - térfogati jellemzői automatikusan kinyerhetők a CAD-adatbázisokból. A CAD-programokban létrehozott parametrikus objektumokat a geometriamag segítségével a hossz, szélesség, terület, térfogat és egyéb paraméterek numerikus értékeivé alakítják át. A 3D geometria alapján történő térfogatnyerés folyamatát részletesen a következő, hatodik részben (6.3-3. ábra) tárgyaljuk, amely a CAD-ekkel való munkának (BIM) van szentelve. A térfogatok mellett a hasonló elemek száma is kinyerhető a CAD-modell adatbázisából az objektumok kategóriák és tulajdonságok szerinti szűrésével és csoportosításával. Ezek a paraméterek, amelyek lehetővé teszik a csoportosítást, a projektelemek összekapcsolásának alapjává válnak az erőforrás-számításokon keresztül a teljes projekt számításaihoz, végső becsléseihez és költségvetéséhez.

Így a 3D (CAD) modellből kinyert adatmodell új paraméteregekkel egészül ki, amelyeket 4D és 5D néven jelölünk. Az új, 4D (idő) és 5D (költség) entitásattribútum rétegekben a 3D geometriai adatok szolgálnak az entitás térfogat attribútumértékek forrásaként.

- **A 4D** egy paraméterinformációs réteg, amely az elemek 3D paramétereihez az építési műveletek időtartamára vonatkozó információkat ad hozzá. Ezek az adatok nélkülözhetetlenek az ütemtervezéshez és a projekt ütemterveinek kezeléséhez
- **Az 5D** az adatmodell következő szintű kiterjesztése, amelyben az elemek költségjellemzőkkal egészülnek ki. Ez a geometriai információkhöz pénzügyi szempontot ad hozzá: az anyagok, munkák és berendezések költségét, ami lehetővé teszi a költségvetési számításokat, a jövedelmezőségi elemzéseket és a költséggazdálkodást az építési folyamat során.

A projektelemek csoporthainak költség- és 3D-, 4D és 5D attribútumadatait a moduláris ERP, PIMS - rendszerek (vagy Excel -szerű eszközök) számításaihoz hasonló módon írják le, és minden az egyes csoportok, minden a teljes projektköltségvetés automatikus költségszámításához és költségvetés-tervezéséhez használják.

5D attribútumok és az attribútum térfogatának kinyerése a CADról

Egy építési projekt végleges becslésének elkészítésekor, amelynek elkészítését az előző fejezetekben már tárgyaltuk (5.1-8. ábra), a projektelemek egyes kategóriáinak térfogati attribútumait vagy kézzel gyűjtik, vagy a CAD szoftverek által megadott térfogati attribútum specifikációkból vonják ki.

A mennyiségek kiszámításának hagyományos kézi módszere a művezető és a becsüs olyan rajzok elemzését jelenti, amelyeket évezredek óta papírra vetett vonalak formájában, az utóbbi 30 évben pedig digitális formátumban, például PDF (PLT) vagy DWG. Szakmai tapasztalataikra támaszkodva, gyakran vonalzóval és szögmérővel mérik ki a szükséges munka- és anyagmennyiségeket. Ez a módszer jelentős erőfeszítést és időt, valamint különös figyelmet igényel.

A terjedelmi jellemzők ilyen módon történő meghatározása a projekt terjedelmétől függően néhány naptól akár több hónapig is eltarthat. Ezenkívül, mivel minden mérést és számítást kézzel végeznek, fennáll az emberi hiba kockázata, ami pontatlan adatokhoz vezethet, ami később kihat a projektidő és -költség becslésének hibáira, amiért az egész vállalatot felelősségre vonják.

A CAD-adatbázisok használatán alapuló modern módszerek nagymértékben leegyszerűsítik a térfogatszámítást. A CAD-modellekben az elemek geometriája már tartalmazza a térfogati attribútumokat, amelyek automatikusan kiszámíthatók (a geometria magon keresztül (6.3-3. ábra)) és táblázatos formában bemutathatók vagy exportálhatók.

Ilyen esetben a becslési osztály a CAD-tervezőtől adatokat kér a projektelemek mennyiségi és térfogati jellemzőiről. Ezeket az adatokat táblázatok formájában exportálják, vagy közvetlenül integrálják a költségszámítási adatbázisokba - legyen az Excel, ERP vagy PMIS - rendszerekbe. Ez a folyamat gyakran nem hivatalos kéréssel kezdődik, hanem az ügyfél (kezdeményező) és az építésza kivitelező vagy tervező cég kalkulátora közötti rövid párbeszéddel. Az alábbiakban egy egyszerűsített példa mutatja be, hogyan alakul ki a minden nap kommunikációból egy strukturált táblázat az automatikus számításokhoz (QTO):

- ❶ Ügyfél - "Szeretnék egy újabb emeletet hozzáadni az épülethez, ugyanabban a konfigurációban, mint az első emelet"
- ❷ Építész (CAD) - "Egy harmadik emelet hozzáadása, a konfiguráció ugyanaz, mint a második emeleten".
És ezután az üzenet után elküldi a projekt új CAD-verzióját a becslőnek.
- ❸ A becslő automatikusan elvégzi a csoportosítást és a számítást (ERP, PMIS, Excel) - "Lefuttatom a projektet egy Excel táblázaton keresztül a QTO szabályokkal (ERP, PMIS), megkapom a mennyiségeket kategóriánként az új emeletre, és generálok egy becslést"

Ennek eredményeképpen a szöveges párbeszédet csoportosítási szabályokkal ellátott táblázatszerkezetté alakítja át:

Elem	Kategória	Padló
Átfedés	OST_Floors	3
Oszlop	OST_StructuralColumns	3
Iépcsőház	OST_Stairs	3

A tervezőtől származó CAD-modell automatikus csoportosítása a becslő QTO szabályai szerint, valamint a mennyiségek automatikus megszorzása az erőforrás-számításokkal (5.1-8. ábra) után a következő eredményeket kapjuk, amelyeket elküldünk a megrendelőnek:

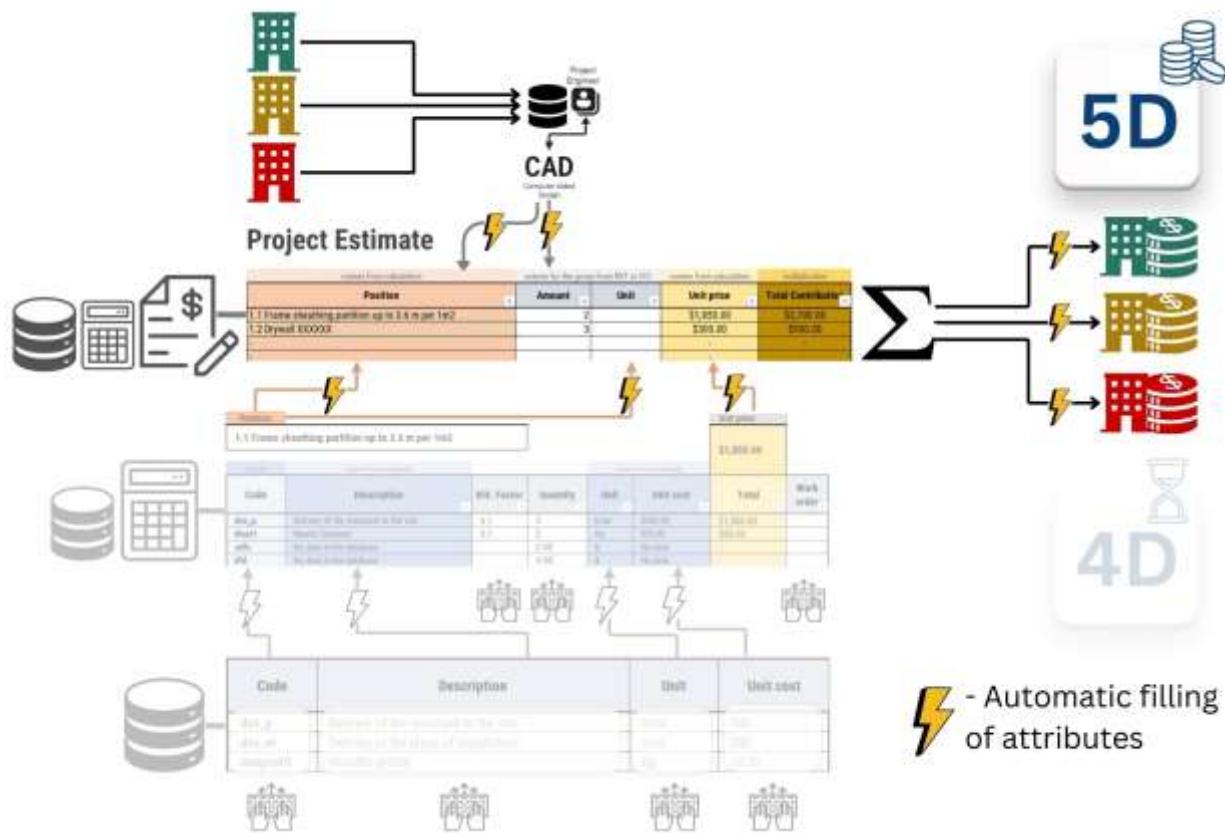
Elem	Kötet	Padló	Ár egységenként.	Teljes költség
Átfedés	420 M ²	3	150 €/M ²	63 000 €
Oszlop	4 db.	3	2450 €/darab.	9 800 €
Iépcsőház	2 db.	3	4 300 €/darab.	8 600 €
ÖSSZESEN:	-	-	-	81 400 €

✉ Az ügyfél - "Köszönöm, ez elég sok, le kell vágnunk néhány szobát". És a ciklus megismétlődik.

Ez a forgatókönyv sokszor megismétlődhet, különösen a jóváhagyási fázisban, ahol az ügyfél azonnali visszajelzést vár. A gyakorlatban azonban az ilyen folyamatok napokig vagy akár hetekig is elhúzódhatnak. Ma már az automatikus csoportosítási és számítási szabályok vezetésének köszönhetően a korábban jelentős időt igénylő tevékenységek percek alatt elvégezhetők. A mennyiségek csoportosítási szabályokon keresztül történő automatizált beszerzése nemcsak felgyorsítja a számításokat és becsléseket, hanem az emberi tényező minimalizálásával csökkenti a hibák valószínűségét, átlátható és pontos projektköltség-értékelést biztosítva.

Ha a CAD-rendszerben a 3D-modell létrehozásakor a becslési osztály követelményeit már eleve figyelembe vették (ami a gyakorlatban még mindig ritka), és az elemcsoportok neveit, azonosítóit és osztályozási attribútumait olyan paraméterek formájában állították be, amelyek egybeesnek a becslési csoportok és osztályok struktúráival, akkor a térfogati attribútumok automatikusan, további transzformációk nélkül átvihetők a becslési rendszerekbe.

A térfogati attribútumok automatikus kivonása a CAD-ból specifikációs táblázatok formájában lehetővé teszi, hogy gyorsan naprakész adatokhoz jussunk az egyes munkák és a projekt egészének költségeiről (5.2-1. ábra). Ha a számítási folyamatban vagy a számítási rendszerben csak a CAD-fájlt frissíti a projekt térfogatával, a vállalat gyorsan újraszámíthatja a becslést a legújabb változások figyelembevételével, biztosítva az összes későbbi számítás nagy pontosságát és konziszenciáját.



Ábra 5.2-1 A CAD-táblázatokból vagy adatbázisokból származó térfogati attribútumok automatikusan bekerülnek a becslésbe, így azonnal kiszámíthatja a projekt teljes költségét.

A tőkeprojektek egyre összetettebbé válásával a teljes költségvetés kiszámítása és a projektek teljes költségének elemzése egy ilyen forgatókönyv szerint (5.2-1. ábra) - a megalapozott döntéshozatal kulcsfontosságú eszközévé válik.

Az Accenture Creating More Value through Capital Projects (2024) című tanulmánya [20] szerint a vezető vállalatok aktívan integrálják az adatelemzést a digitális kezdeményezésekbe, és a múltbeli információkat felhasználják az eredmények előrejelzésére és optimalizálására. A kutatás azt mutatja, hogy egyre több tulajdonos-üzemeltető alkalmazza a nagy adatelemzést a piaci trendek előrejelzésére és a kereskedelmi életképesség értékelésére, még a tervezés megkezdése előtt. Ezt a meglévő projektporfolióból származó adattárházak elemzésével érik el. Emellett a tulajdonos-üzemeltetők 79 százaléka "robusztus" előrejelző elemzést alkalmaz a projektek teljesítményének értékelésére és a valós idejű operatív döntéshozatal támogatására.

Az építési projektek modern, hatékony irányítása elválaszthatatlanul kapcsolódik a nagy mennyiségű információ feldolgozásához és elemzéséhez a tervezés és a tervezést megelőző folyamatok minden szakaszában. A adattárházak, az erőforrás-számítások, az előrejelző modellek és a gépi tanulás használata nemcsak a számítások kockázatainak minimalizálását teszi lehetővé, hanem azt is, hogy a tervezés korai szakaszában stratégiai döntéseket hozzanak a projektfinanszírozásról. Az adattárházakról és a számításokat kiegészítő prediktív modellek ról a könyv kilencedik részében fogunk bővebben beszélni.

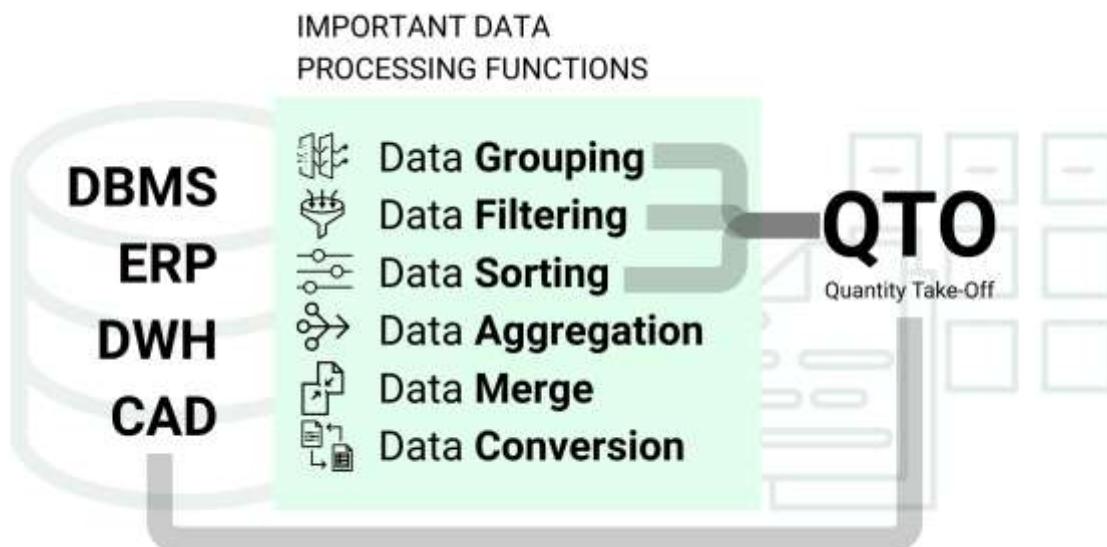
A becslések elkészítéséhez szükséges elemek térfogati paramétereinek automatikus kinyerése a CAD-

projektekből a QTO (Quantity Take-Off) csoportosító eszközök segítségével történik. A QTO-eszközök úgy működnek, hogy a CAD-adatbázisban létrehozott specifikációk és táblázatok segítségével az összes projektobjektumot speciális elemazonosítók vagy elemattribútum-paraméterek alapján csoportosítják.

QTO Quantity Take-Off: projektadatok csoportosítása attribútumok szerint

A QTO (Quantity Take-Off) az építőiparban a projekt megvalósításához szükséges elemek mennyiségeinek meghatározása. A gyakorlatban a QTO gyakran félre manuális folyamat, amely magában foglalja az adatok gyűjtését különböző forrásokból: PDF dokumentumokból, DWG rajzokból és digitális CAD modellekben.

A CAD-adatbázisokból kinyert adatokkal való munka során a QTO-folyamat szűrési, rendezési, csoportosítási és aggregálási műveletek sorozataként valósul meg. A modellelemek kiválasztása az osztályok, kategóriák és típusok paraméterei szerint történik, majd mennyiségi attribútumaik - például térfogat, terület, hossz vagy mennyiség - a számítási logika szerint összegzésre kerülnek (5.2-2. ábra).



5.2-2 Az adatok csoportosítása és szűrése az adatbázisokban és adattárházakban alkalmazott legnépszerűbb funkciók.

A QTO (szűrés és csoportosítás) folyamat lehetővé teszi az adatok rendszerezését, a specifikációk kialakítását és a becslések, beszerzések és munkatervezek kiszámításához szükséges bemeneti információk előkészítését. A QTO alapja az elemek osztályozása a mért jellemzők típusa szerint. minden egyes elemhez vagy elemcsoporthoz kiválasztják a megfelelő mennyiségi mérési paramétert. Például:

- **Hosszúsági attribútum** (szegélykő - méterben)
- **Területjellemző** (gipszkartonozás - négyzetméterben)
- **Térfogati attribútum** (betonművek - köbméterben)
- **Mennyiségi attribútum** (ablakok - darabonként)

A geometria alapján matematikailag előállított térfogati jellemzők mellett a QTO-csoportosítás utáni számításokban gyakran alkalmaznak túllépési tényezőket (5.2-12 ábra, pl. 1,1 a logisztika és a telepítés 10

százalékának figyelembevételére) - olyan korrekciós értékeket, amelyek figyelembe veszik a veszteségeket, a telepítést, a tárolást vagy a szállítást. Ez lehetővé teszi az anyagok tényleges fogyasztásának pontosabb előrejelzését, és az építkezésen minden hiányt, minden túlkínálat elkerülését.

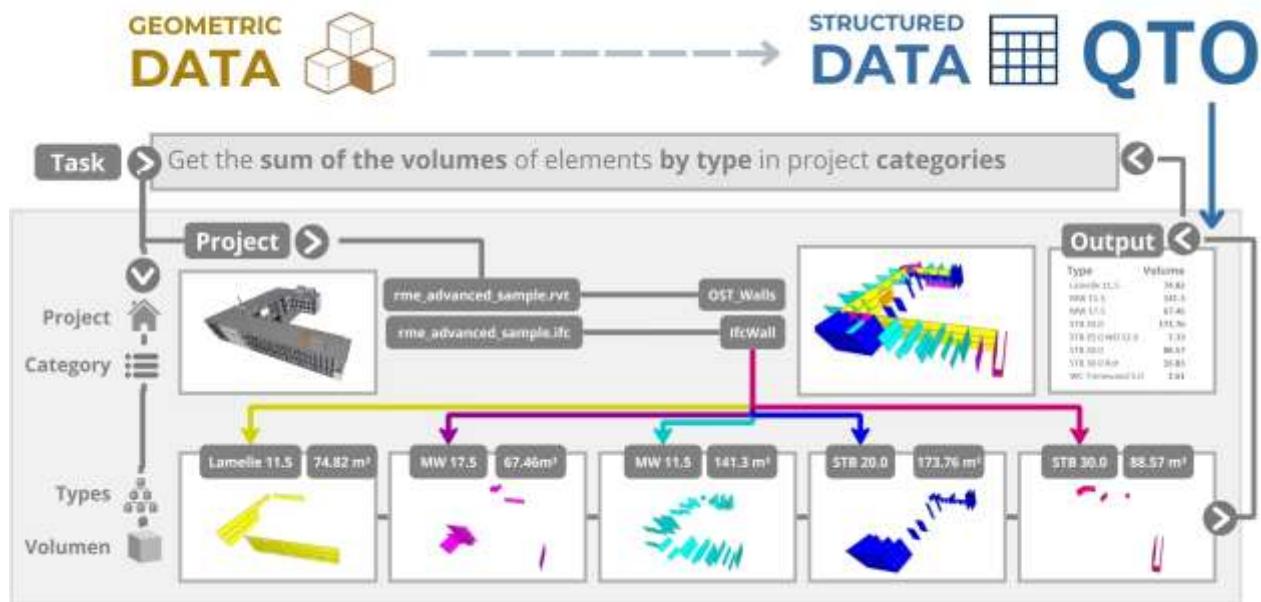
Az automatizált mennyiségi átvételi folyamat (QTO) elengedhetetlen a pontos számítások és becslések elkészítéséhez, a mennyiségi specifikációs folyamatok emberi hibáinak csökkentéséhez és az anyagok túl- vagy alulrendelésének megelőzéséhez.

A QTO folyamat példájaként tekintsünk egy gyakori esetet, amikor a CAD-adatbázisból meg kell jeleníteni a térfogatok táblázatos specifikációját elemtípusok szerint egy bizonyos kategóriára, elemesztályokra vonatkozóan. Csoportosítuk az összes projektelemet típusok szerint a CAD-projektfal kategóriájából, és foglaljuk össze az egyes típusok térfogatattribútumait, hogy az eredményt QTO térfogattáblázat formájában mutassuk be (5.2-3. ábra).

Egy tipikus CAD-projekt példájában (5.2-3. ábra) a CAD-adatbázisban található összes falkategóriájú elem a fal típusa szerint van csoportosítva, pl. "Lamelle 11.5", "MW 11.5" és "STB 20.0", és jól meghatározott térfogati attribútumokkal rendelkezik, amelyeket metrikus kockákban ábrázolnak.

A tervezők és a számítási szakemberek közötti kapcsolódási ponton álló menedzser célja, hogy a kiválasztott kategóriában az elemtípusok szerinti térfogatokat tartalmazó automatikus táblázatot kapjon. Nemcsak egy adott projektre vonatkozóan, hanem egy univerzális formában is, amely más, hasonló modellszerkezetű projektekre is alkalmazható. Ez lehetővé teszi a megközelítés skálázhatóságát és az adatok újrafelhasználását a munka megkettőzése nélkül.

Elmúltak azok az idők, amikor a tapasztalt tervezők és a becslők vonalzóval felfegyverkezve, gondosan bemérve minden egyes vonalat a papíron vagy a PDF -terveken - ez a hagyomány nem változott az elmúlt évezredek során. A 3D -modellezés fejlődésével, ahol az egyes elemek geometriája ma már közvetlenül kapcsolódik az automatikusan kiszámított térfogati jellemzőkhöz, a térfogatok és a QTO mennyiségek meghatározásának folyamata automatizálttá vált.



5.2-3 ábra A QTO mennyiségi és mennyiségi attribútumok kinyerése egy projektből a projektelemek csoportosítását és szűrését jelenti.

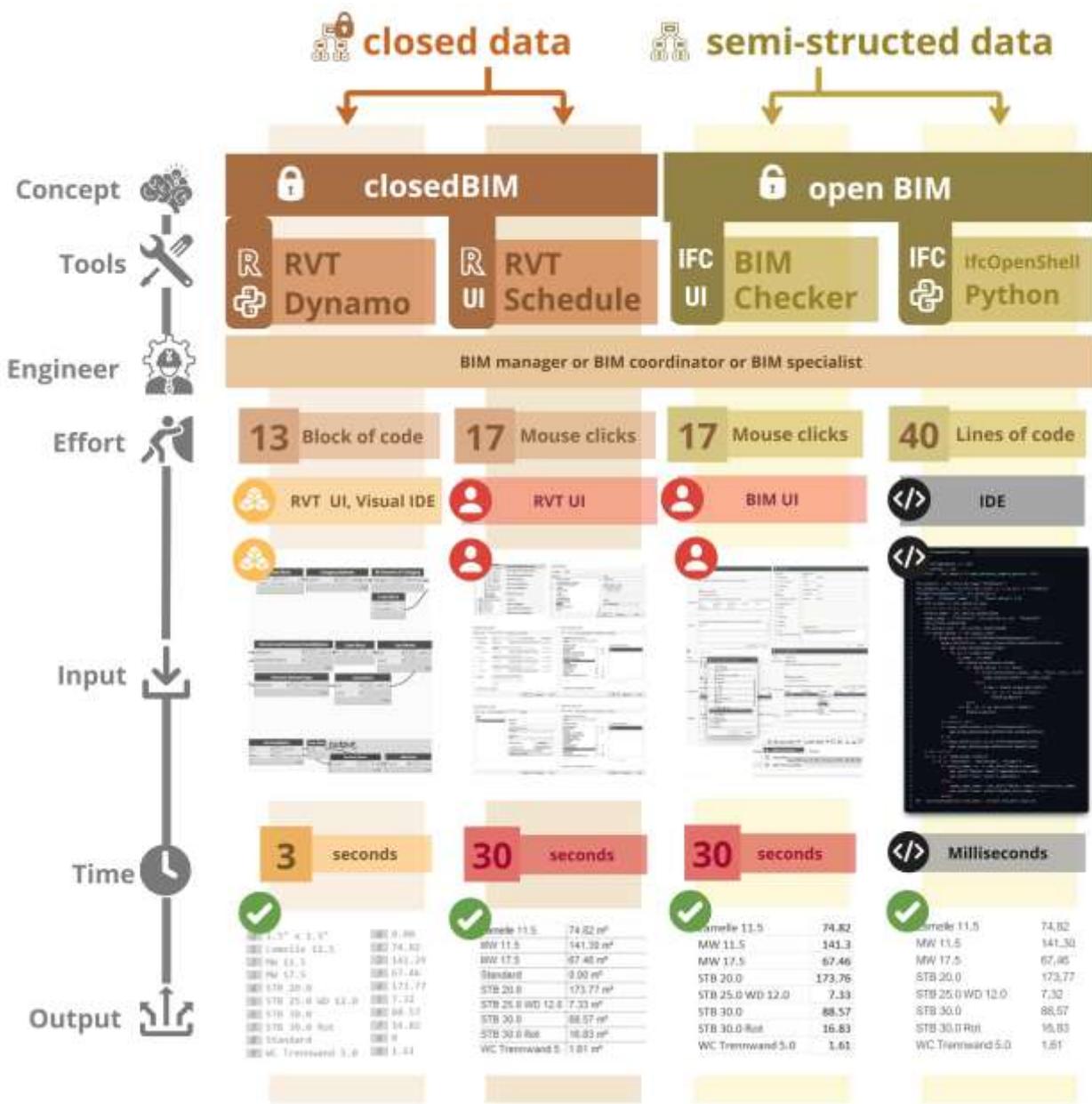
Példánkban a feladat "egy projektben a falak egy kategóriájának kiválasztása, az összes elem típus szerinti csoportosítása és a mennyiségi jellemzőkkel kapcsolatos információk strukturált, táblázatos formában történő bemutatása", hogy ezt a táblázatot több tucat más szakember használhassa a költségszámítások, logisztika, munkarendek és egyéb üzleti esetek számításaihoz (6.1-3. ábra).

A CAD-adatok zárt jellege miatt ma már nem minden szakember használhatja a CAD-adatbázishoz való közvetlen hozzáférést (a hozzáférési probléma okait és megoldásait a könyv hatodik része részletezi). Ezért sokan kénytelenek a nyílt BIM és a zárt BIM [63] koncepcióin alapuló speciális BIM-eszközökhöz fordulni. A specializált BIM eszközökkel vagy közvetlenül a CAD szoftverkörnyezetben történő munka során a QTO (Quantity Take-Off) eredményeket tartalmazó táblázatot különböző módon lehet létrehozni - attól függően, hogy kézi felületet vagy szoftveres automatizálást használunk.

Például a CAD (BIM) szoftverek felhasználói felületét használva elegendő körülbelül 17 műveletet (gombnyomás) végrehajtani ahhoz, hogy egy kész térfogattáblázatot kapunk (5.2-4. ábra). A felhasználónak azonban jól kell ismernie a CAD (BIM) szoftver modellszerkezetét és funkciót.

Ha az automatizálást programkódon vagy a CAD-programokon belüli beépülő modulokon és API-eszközökön keresztül alkalmazzuk, a térfogattáblák előállításához szükséges kézi lépések száma csökken, de a használt könyvtár vagy eszköz függvényében 40-150 sornyi kódot kell írni:

- **IfcOpSh (nyílt BIM)** vagy **Dynamo IronPython (zárt BIM)** - lehetővé teszi, hogy egy CAD formátumból vagy CAD programból egy QTO táblázatot kapjon minden össze ~40 sornyi kóddal.
- **IFC_js (nyílt BIM)** - körülbelül 150 sornyi kódot igényel a terjedelmes attribútumok kinyeréséhez az IFC modellből.
- **Interface CAD eszközök (BIM)** - lehetővé teszi, hogy ugyanazt az eredményt kézzel, 17 egérkattintással érje el.



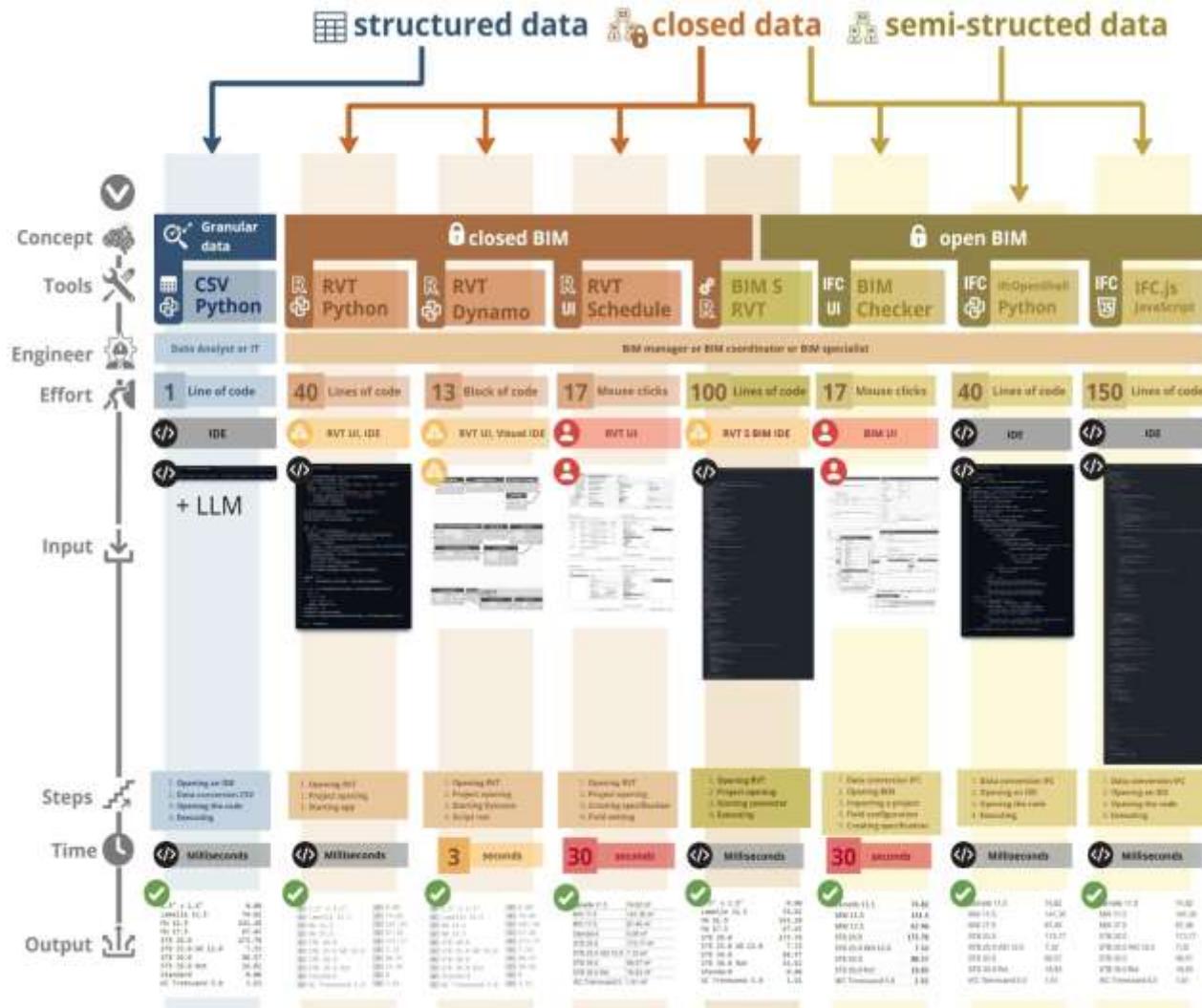
Ábra 5.2-4 CAD (BIM) tervezők és menedzserek 40-150 sor kódot vagy egy tucat billentyűléjtést használnak a QTO táblázatok létrehozására

Az eredmény ugyanaz - egy strukturált táblázat, amely az elemek egy csoportjának kötetattribútumait tartalmazza. Az egyetlen különbség a munkadíjban és a felhasználó által szükséges technikai szakértelemben rejlik (5.2-4. ábra). A modern eszközök a kötetek kézi gyűjtéséhez képest jelentősen felgyorsítják a QTO folyamatot és csökkentik a hibák valószínűségét. Lehetővé teszik, hogy az adatokat közvetlenül a projektmodellből lehessen kinyerni, így nem kell a térfogatokat kézzel újraszámolni a rajzokból, ahogyan azt korábban tették.

Az alkalmazott módszertől függetlenül - legyen az nyílt BIM vagy zárt BIM - azonos QTO - táblázatot kaphatunk a projektelemek térfogatával (5.2-4. ábra). A CAD - (BIM-) koncepciókban a projektadatokkal

való munka során azonban a gyártók által biztosított speciális eszközökre és API-kra szorulnak (3.2-13. ábra). Ez további függőségi rétegeket hoz létre, és egyedi adatsémák elsajátítását igényli, miközben korlátozza az adatokhoz való közvetlen hozzáférést.

A CAD-adatok zárt jellege miatt a QTO-táblázatok és egyéb paraméterek beszerzése megnehezíti a számítások automatizálását és a külső rendszerekkel való integrációt. A adatbázisokhoz való közvetlen hozzáférést biztosító eszközök használatával, valamint a projekt CAD -adatainak a visszafejtési eszközökkel történő átvitelével egy nyílt, strukturált adatkeret formátumba (4.1-13. ábra) egy azonos QTO táblázatot lehet kapni egyetlen sor kóddal (5.2-5. ábra - változat szemcsés adatokkal).



5.2-5. ábra A különböző eszközök a projektelemek attribútumtáblái formájában ugyanazokat az eredményeket adják, de eltérő munkaköltséggel.

A CAD projektekből származó nyílt strukturált adatok használata esetén, amint azt a "CAD (BIM) adatok strukturált formába történő átalakítása" című fejezetben említettük, a csoporthoz köthető folyamat, a QTO nagymértékben leegyszerűsödik.

A nyílt strukturált adatok használata vagy a CAD-modell-adatbázisokhoz való közvetlen hozzáférésen alapuló megközelítések mentesülnek a BIM rövidítéshez kapcsolódó marketingkényszertől. Más

iparágakban már régóta használt, bevált eszközökre támaszkodnak (7.3-10. ábra ETL folyamat).

A McKinsey "Open Data: Unleash Innovation and Productivity with Streaming Information" (Innováció és termelékenység felszabadítása az információáramlással). [102] 2013-ban készített tanulmánya szerint a nyílt adatok használata évente 30-50 milliárd dollár megtakarítási lehetőséget teremthet az energiatermelő létesítmények tervezése, mérnöki munkái, beszerzése és építése során. Ez az építési tőkeköltségek 15 százalékos megtakarítását jelenti.

A nyílt strukturált (granuláris) adatokkal való munka leegyszerűsíti az információkeresést és -feldolgozást, csökkenti a speciális BIM-platformuktól való függőséget, és megnyitja az utat az automatizálás előtt anélkül, hogy szükség lenne szabadalmaztatott rendszerek vagy CAD-formátumokból származó parametrikus és összetett adatmodellek használatára.

QTO automatizálás az LLM és a strukturált adatok használatával.

A strukturálatlan adatok strukturált formába történő lefordítása jelentősen javítja a különböző folyamatok hatékonyságát: egyszerűsíti az adatfeldolgozást (4.1-1. ábra, 4.1-2. ábra) és felgyorsítja a validálási folyamatot azáltal, hogy a követelményeket egyértelművé és átláthatóvá teszi, amint azt már az előző fejezetekben tárgyaltuk. Hasonlóképpen, a CAD-adatok (BIM) strukturált nyílt formába történő lefordítása (4.1-12. ábra, 4.1-13. ábra) megkönnyíti az attribútumcsoportosítási folyamatot és a QTO folyamatot.

A QTO attribútumtábla strukturált formában van, így a strukturált CAD adatok használatakor egyetlen adatmodellel dolgozunk (5.2-5. ábra), ami kiküszöböli a projekt adatmodellek és csoportosítási szabályok közös nevezőre való konvertálásának és lefordításának szükségességét. Ez lehetővé teszi számunkra, hogy az adatokat egy vagy több attribútum alapján csoportosítsuk egyetlen sor kóddal. Ezzel szemben a nyílt BIM és a zárt BIM esetében, ahol az adatokat félig strukturált, parametrikus vagy zárt formátumban tárolják, a feldolgozáshoz több tucat vagy akár több száz sornyi kódra van szükség, valamint a API használatára a geometria- és attribútuminformációkkal való interakcióhoz.

- ❷ Példa egy QTO strukturált projekt egy attribútum szerinti csoportosítására. Szöveges lekérdezés bármely LLM csevegőszobában (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vagy bármely más):

Van CAD -project mint DataFrame - kérem szűrje a projektadatokat, hogy olyan elemeket kapjon, amelyek "Type" paramétere csak "Type 1" értéket tartalmaz. ↵

- Az LLM válasza nagy valószínűsséggel Python kód formájában Pandas segítségével:



5.2-6 ábra Az LLM segítségével írt egyetlen kód sor lehetővé teszi, hogy egy teljes CAD projektet a "Type" attribútum alapján csoportosítson, és megkapja a kívánt elemcsoportot.

A kétdimenziós DataFrame egyszerű felépítésének köszönhetően nem kell elmagyaráznunk az LLM sémát és adatmodellt, ami lerövidíti az értelmezési lépéseket és felgyorsítja a végleges megoldások létrehozását. Korábban még az egyszerű kód megírásához is programozási nyelvek tanulására volt szükség, de ma már a modern nyelvi modellek (LLM) lehetővé teszik, hogy a strukturált adatokkal való munka során a folyamatlogikát automatikusan kódra fordítsuk le szöveges lekérdezések segítségével.

Az LLM automatizálás és a nyelvi modellek teljesen kiküszöbölhetik, hogy a CAD (BIM) adatcsoportosítással és -feldolgozással foglalkozó szakembereknek programozási nyelveket vagy BIM-eszközöket kelljen tanulniuk azáltal, hogy lehetővé teszik a problémák szöveges lekérdezésekkel történő megoldását.

Ugyanaz a lekérdezés - a "falak" kategóriába tartozó összes projektelem csoportosítása és az egyes típusok térfogatának kiszámítása (5.2-5. ábra) - amely egy CAD-környezetben (BIM) 17 kattintást igényel a felületen, vagy 40 sor kód megírását, nyílt adatfeldolgozó eszközökben (pl. SQL vagy Pandas) egyszerű és intuitív lekérdezésnek tűnik:

- Egyetlen sorral a Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

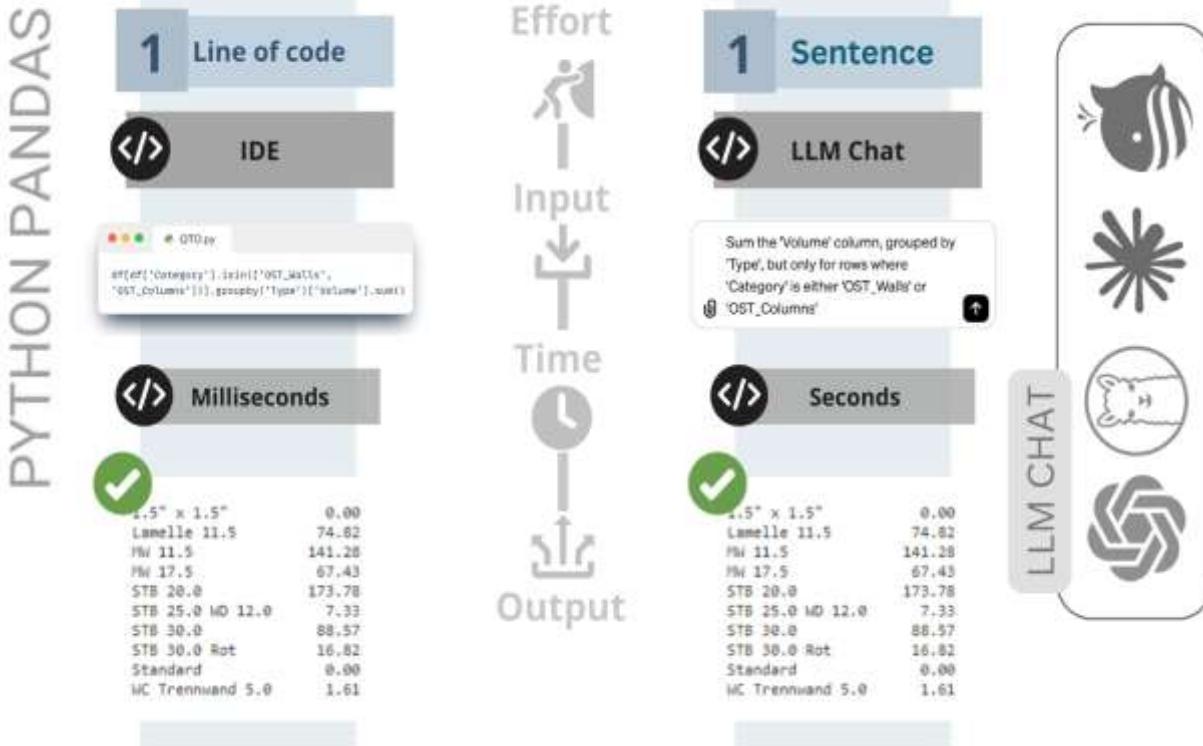
Kódfejtés: vegye ki a df-ből (DataFrame) azokat az elemeket, amelyek "Category" attribútum-oszlopának értékei "OST_Walls", csoportosítsa az összes kapott elemet a "Type" attribútum-oszlop szerint, és összegezze a kapott elemcsoportra a "Volume" attribútumot.

- A CAD-ból származó strukturált projekt csoportosítása SQL segítségével:

```
SELECT Type, SUM(Volume) AS TotalVolume (Teljes mennyiség)
FROM elements
WHERE Kategória = 'OST_Falak'
GROUP BY Típus;
```

- Az LLM segítségével egyszerű szöveges hivatkozásként - promptként - csoportosítási kérést írhatunk a projektadatbázisba (5.2-7. ábra):

A projekt adatkeret esetében csoportosítsa a tételeket a "Típus" paraméter szerint, de csak azon tételek esetében, amelyek "Kategória" paramétere "OST_Falak" vagy "OST_oszlopok", és foglalja össze a "Mennyiség" oszlopparamétert az így kapott vonatkozóan. ↴

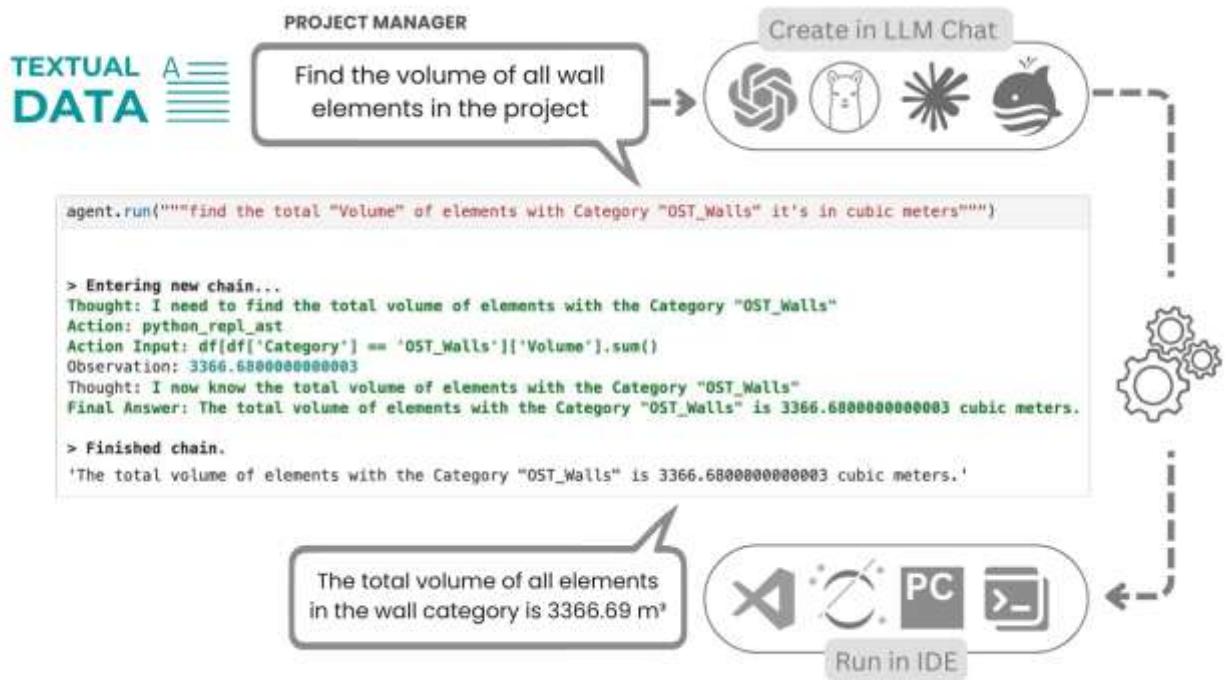


5.2-7 ábra SQL, Pandas és LLM használatával az adatfeldolgozás automatizálása néhány sornyi kódossal és szöveges lekérdezéssel lehetséges.

A QTO kinyerése a CAD-adatokból LLM-eszközök (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) segítségével drámaian megváltoztatja az attribútuminformációk, mennyiségi és volumetrikus adatok kinyerésének hagyományos módszereit az egyes objektumok és objektumcsoportok esetében.

Most még azok a projektmenedzserek, költségszámítási vagy logisztikai szakemberek is, akik nem rendelkeznek mélyreható tervezési ismeretekkel és nem rendelkeznek speciális CAD szoftverrel - (BIM-) gyártók, akik hozzáférnek a CAD-adatbázishoz, másodpercek alatt, egyszerűen egy lekérdezés megírásával vagy diktálásával megkaphatják a falak vagy más objektumok kategóriájába tartozó elemek teljes mennyiségett.

A szöveges lekérdezéseknel (5.2-8. ábra) a modell LLM-ügynöke feldolgozza a felhasználó kérést, hogy egy bizonyos függvényt alkalmazzon egy vagy több paraméterre - a táblázat oszlopaira. Ennek eredményeként az LLM-mel kommunikáló felhasználó vagy egy új oszlop-paramétert kap új értékekkel, vagy egy adott értéket a csoportosítás után



Ábra 5.2-8 LLM modell, amely strukturált adatokkal dolgozik, a szöveges lekérdezés kontextusából megérti, hogy a felhasználó milyen csoportosításra és attribútumokra kérdez rá.

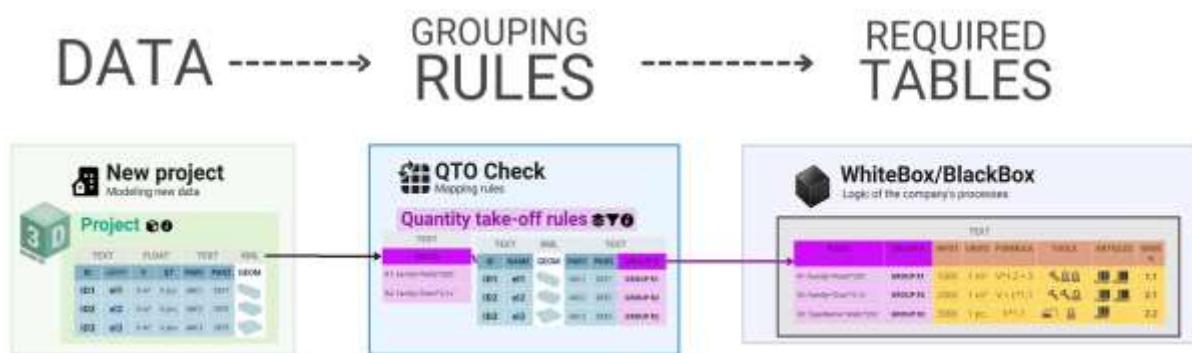
Ha csak egy elemcsoportra vonatkozó mennyiségeket kell beszerezni, elegendő egy egyszerű QTO-lekérdezést (5.2-7. ábra) végrehajtani a CAD-modell adatain. Ha azonban egy teljes, sok elemcsoportból álló projekt költségvetésének vagy becslésének kiszámításakor gyakran szükséges a mennyiségek kinyerése az összes elemtípusra (osztályokra), ahol minden egyes elemkategóriát külön-külön kell feldolgozni - a vonatkozó attribútumok szerinti csoportosítással.

A becslők és értékelők gyakorlatában a különböző típusú tárgyakra egyedi csoportosítási és számítási szabályokat alkalmaznak. Például az ablakokat általában emeletek vagy zónák szerint csoportosítják (csoportosítási paraméter - attribútum Szint, helyiségek), a falakat pedig anyag vagy építési típus szerint (paraméter Anyag, típus). A csoportosítási folyamat automatizálása érdekében az ilyen szabályokat előre le kell írni a oldalon csoportosítási szabály táblázatok formájában. Ezek a táblázatok konfigurációs

sablonokként működnek, amelyek meghatározzák, hogy a projekt egyes elemcsoportjainak számításában mely attribútumokat kell használni.

QTO A teljes projekt számítása Excel táblázatból származó csoportszabályok segítségével.

A valós építési projektekben gyakran előfordul, hogy egy elemcsoporton belül egyszerre több attribútum szerinti aggregációt kell végrehajtani. Például az "Ablakok" kategóriával való munka során (ahol a Kategória attribútum olyan értékeket tartalmaz, mint az OST_Ablakok vagy az IfcAblakok), az elemek nemcsak típus szerint csoportosíthatók - például a Típusnév vagy a Típus mező értéke alapján -, hanem további jellemzők, például a megfelelő attribútumban megadott hővezetési szint alapján is. Ez a többdimenziós csoportosítás pontosabb eredményeket tesz lehetővé egy adott csoportra vonatkozóan. Hasonlóképpen, a fal- vagy födémkategóriák kiszámításakor az attribútumok - például az anyag, a szint, a padló, a tűzállóság és más paraméterek - tetszőleges kombinációi szűrőként vagy csoportosítási kritériumként használhatók (5.2-9. ábra).



5.2-9 ábra A projektben szereplő entitások minden egyes csoportjára vagy kategóriájára egy vagy több kritériumból álló, különböző csoportosítási képlet létezik.

Az ilyen csoportosítási szabályok meghatározásának folyamata hasonló az adatkövetelmények létrehozásának folyamatához, amelyet a "Követelmények létrehozása és minőségellenőrzés adatok" című fejezetben (4.4-5. ábra) ismertünk, ahol részletesen tárgyaltuk az adatmodellekkel való munkát. Az ilyen csoportosítási és számítási szabályok biztosítják az eredmények pontosságát és relevanciáját, hogy automatikusan kiszámíthassuk egy entitáskategória mennyiségeinek vagy mennyiségeinek összes attribútumát, figyelembe véve minden olyan szükséges feltételt, amelyet a számítások és számítások során figyelembe kell venni.

- ➲ A következő kód minta úgy szűri a projektek táblázatot, hogy az eredményül kapott adathalmaz csak olyan entitásokat tartalmazzon, amelyeknél a "Kategória" attribútum-oszlop az "OST_Windows" vagy az "IfcWindows" értékét tartalmazza, és ugyanakkor a "Típus" attribútum-oszlop az "1. típus" értékét tartalmazza:

Van egy DataFrame projektem - szűrje az adatokat úgy, hogy csak olyan elemek maradjanak az adatkészletben, amelyeknek a "Category" attribútuma az "OST_Windows" vagy "IfcWindows" értékeket tartalmazza, és ugyanakkor a Type attribútum az "Type 1" értéket tartalmazza. ↴

▣ LLM válasza :

```
group.py
df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

5.2-10 ábra A Excel-képlethez hasonlóan egyetlen kódsor segítségével az összes projektelemet több attribútum alapján csoportosíthatja.

Az így kapott kód (5.2-10. ábra) a CAD-adatok strukturált nyílt formátumban történő lefordítása után (4. ábra).1-13) futtatható a fent említett népszerű IDE-k (integrált fejlesztőkörnyezetek) valamelyikében offline üzemmódban: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev pluginnal, Thonny, Wing IDE, IntelliJ IDEA Python pluginnal, JupyterLab vagy népszerű online eszközök: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- ▣ A QTO DataFrame űrlapon a "Windows" kategória alatt található projekt entitások lekérdezéséhez csak egy adott hővezetési értékkel, a következő lekérdezéssel kereshetjük le az LLM-et:

Van egy DataFrame projektem - szűrje az adatokat úgy, hogy csak az "OST_Windows" vagy "IfcWindows" értékeket tartalmazó "Category" rekordok maradjanak az adatkészletben, és ugyanakkor a ThermalConductivity oszlopnak 0. értékűnek kell lennie.
↳

💡 LLM válasza :



The screenshot shows a user interface for generating code in an LLM chat. At the top right is a button labeled "Create in LLM Chat". Below it are four icons: a blue circle with a white swirl, a white circle with a blue square, a yellow starburst, and a white whale. A code editor window titled "group.py" contains the following Python code:

```
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

Below the code editor are several execution options: "Run in iDE" (with icons for VS Code, PyCharm, and Eclipse), "Run in Jupyter Notebook" (with a Jupyter logo), "Run in Colab" (with a Google Colab logo), and "Run in Terminal" (with a terminal icon).

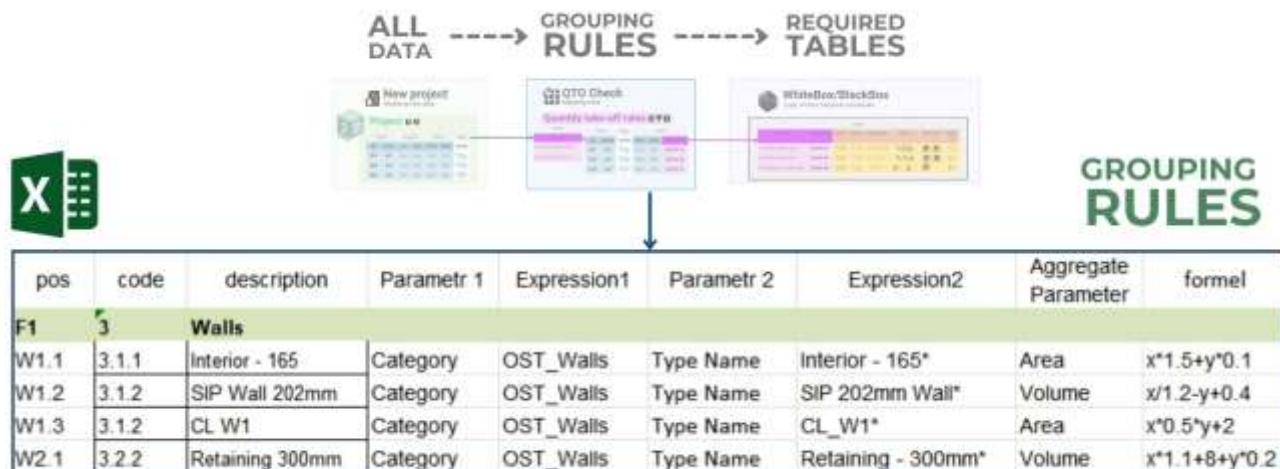
5.2-11 ábra A rendkívül egyszerű Pandas lekérdezési nyelv Python lehetővé teszi, hogy a QTO-kat tetszőleges számú projektre egyidejűleg futtassa.

Az LLM-től kapott válaszban (5.2-11. ábra) az "&" logikai feltétel két kritérium kombinálására szolgál: a hővezetési érték és a két kategória egyikébe való tartozás. Az "isin" módszer ellenőrzi, hogy a "Kategória" attribútum oszlop értéke szerepel-e a megadott listában.

A nagyszámú elemcsoportot tartalmazó, eltérő csoportosítási logikával rendelkező projektek esetében - a projektelemek minden egyes kategóriájára (pl.: ablakok, ajtók, födémek) egyedi csoportosítási szabályokat kell megállapítani, amelyek tartalmazhatnak további együtthatókat vagy az összes attribútum számítási képleteit. Ezek a képletek (5.2-12. ábra attribútum "formel", pl. a mennyiség x-értéke és a csoport y-terfogata) és együtthatók figyelembe veszik például az egyes csoportok egyedi jellemzőit:

- az anyagmennyiséghoz a túllépések figyelembevétele érdekében tett%-os kiegészítések
- rögzített kiegészítő anyagmennyiség
- a képletekben szereplő lehetséges kockázatokkal és számítási hibákkal kapcsolatos kiigazítások

Miután a szűrési és csoportosítási szabályokat paraméteres képletek formájában minden egyes térelkategóriára vonatkozóan megfogalmaztuk, ezeket soronként táblázat formájában tárolhatjuk - például Excel formátumban (5.2-12. ábra). E szabályok strukturált formában történő tárolásával a projektadatok kinyerésének, szűrésének és csoportosításának folyamata teljesen automatizálható. A sok különálló lekérdezés kézi megírása helyett a rendszer egyszerűen beolvassa a paramétertáblát, és a megfelelő szabályokat alkalmazza a modellre (a teljes projekt adatkeretre (4.1-13. ábra)), létrehozva a végleges QTO táblázatokat a projektelemek minden egyes kategóriájára.



5.2-12. ábra QTO Attribútum csoportosítása A táblázat a projektelemek csoportosítására vonatkozó szabályokat állapítja meg, pontos összlétszámot és terjedelmet biztosítva az egyes kategóriákhoz.

Az összegyűjtött szabályok lehetővé teszik a teljes projekt csoportosítását és az összes szükséges számítás elvégzését, beleértve a térfogati attribútumok kiigazítását is. Ennek eredményeképpen a kötetek a "valós kötetbe" kerülnek, amelyet a számításokhoz és a számításokhoz használnak, és nem azok, amelyek eredetileg a tervezési szakaszban voltak a CAD-modellben.

A teljes projektre vonatkozó QTO volumentáblák automatikus létrehozása során az alkalmazásnak végig kell mennie a csoportosítási szabályok táblázat összes kategóriáján, ki kell vennie a csoportosítási attribútumokat, csoportosítania kell az összes projektelemet ezek alapján, és aggregálnia kell az adott csoportra vonatkozó volumenattribútumot, amelyet ráadásul meg kell szoroznia egy finomító tényezővel vagy együtthatóval.

Kérjük meg az LLM-ét, hogy írjon számunkra kódot egy ilyen megoldásra, ahol a kódnak két táblázatot kellene betöltenie - egy csoportosítási szabályokat tartalmazó táblázatot (5.2-12. ábra) és egy táblázatot magából a projektből származó adatokkal (4.1-13. ábra), majd alkalmazni a csoportosítási szabályokat, csoportosítani az elemeket a megadott szabályok szerint, kiszámítani az összesített értékeket és az eredményeket egy új Excel fájlba menteni.

✉ Küldj egy szöveges kérést az LLM chatre:

Szükségem van egy kódra, amely beolvassa a projektadatokat a 'basic_sample_project.xlsx' fájlból, majd a 'Grouping_rules_QTO.xlsx' fájlból származó szabályokat, amelyekkel az összes adatot az '1. paraméter' és '2. paraméter' szerint csoportosítja, az 'Aggregate Parameter', a 'Expression2' szerint szűr, számításokat végez a 'Formel1'-ből, és elmenti a QTO táblázatot a 'QTQ_table2.xlsx' fájlbba. ↴

☒ LLM válasza :



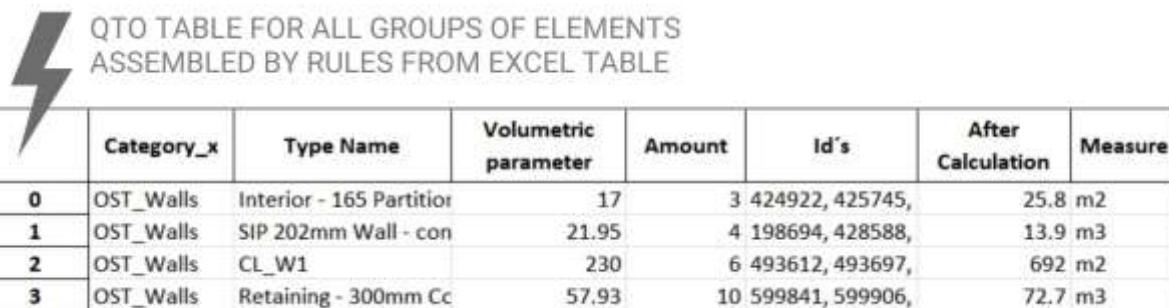
```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 ].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.\d*)')[0], errors='coerce'))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     ][row["Aggregate Parameter"]].agg(['sum', 'count'])
17     group = group[group.index.get_level_values(1).str.contains(row["Expression2"])]
18
19     # Perform calculation defined in the rules
20     group['After_Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
21
22     return group.reset_index()
23
24 # Apply the function to each row in the grouping rules dataframe and concatenate the results
25 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()])
26
27 # Save the processed data to an Excel file using the OpenPyXL engine
28 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

5.2-13. ábra Python kód Pandas feldolgozza a CAD projektből származó tervezési adatokat, és az Excelben meghatározott szabályok szerint csoportosítja azokat .

A kódvégrehajtás végeredménye (5.2-13. ábra) egy olyan entitáscsoport-tábla lesz, amely nemcsak az eredeti CAD- (BIM-) modell általánosított térfogatattribútumait tartalmazza, hanem egy új valós térfogatattribútumot is, amely figyelembe veszi az összes követelményt a számítások és becslések helyes elkészítéséhez (példa: 5.2-14. ábra).

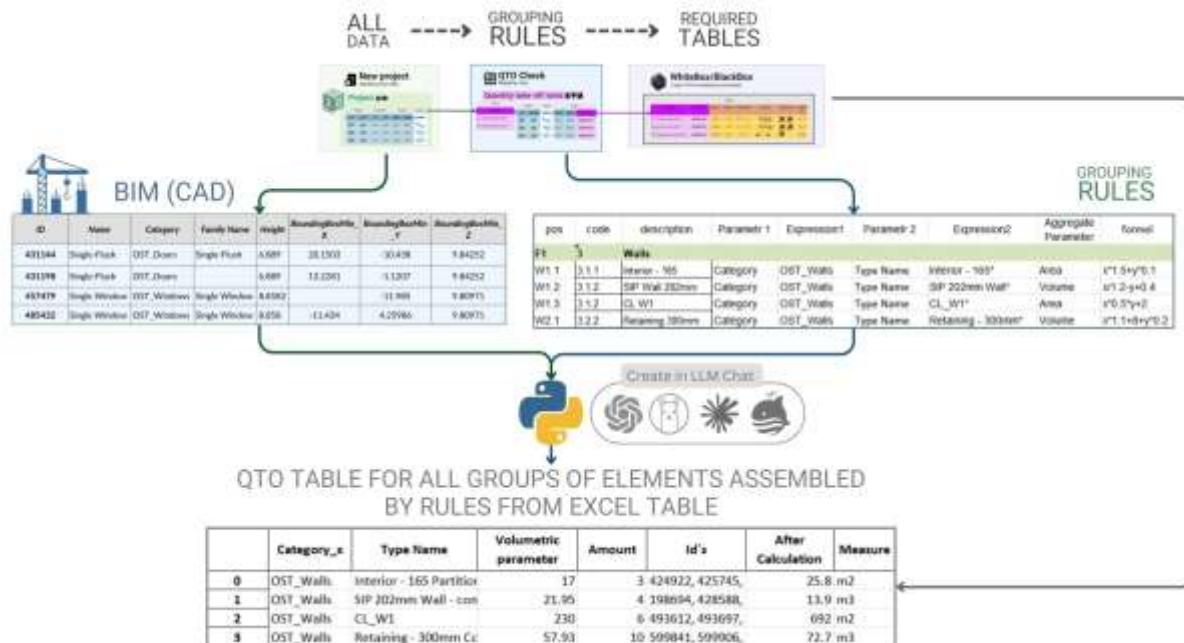


	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition	17	3 424922, 425745,		25.8 m ²	
1	OST_Walls	SIP 202mm Wall - con	21.95	4 198694, 428588,		13.9 m ³	
2	OST_Walls	CL_W1	230	6 493612, 493697,		692 m ²	
3	OST_Walls	Retaining - 300mm Cc	57.93	10 599841, 599906,		72.7 m ³	

ábra 5.2-14 Az "After Calculation" attribútum az összesítő táblázatba kerül, miután a kódot végrehajtották, amely automatikusan kiszámítja a tényleges mennyiséget.

Az így kapott kód (5.2-13. ábra) futtatható valamelyik népszerű IDE-ben (amit fentebb már említettünk), és

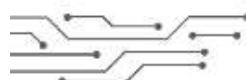
tetszőleges számú meglévő vagy újonnan bejövő projektre (RVT, IFC, DWG, NWS, DGN stb.) alkalmazható, legyen az néhány projekt vagy esetleg több száz különböző formátumú projekt strukturált formában (5.2-15. ábra).



Ábra 5.2-15 Az automatikus építési adatok csoportosítási folyamata a BIM adatokat (CAD) és a QTO táblázatok összekapcsolja egy Excel táblázatból származó szabályok segítségével.

A testreszabott és paraméterezett volumetrikus adatgyűjtési folyamat (5.2-15. ábra) lehetővé teszi a projektelemek mennyiségi jellemzőire és volumenére vonatkozó adatok teljesen automatizált gyűjtését a velük való további munkához, beleérve a költségbecsleést, a logisztikát, a munkaprogramokat, a szénlábnyom-számítást és egyéb elemzési feladatokat.

Miután megtanultuk azokat az eszközöket, amelyek lehetővé teszik számunkra, hogy a projektelemek csoportjait bizonyos attribútumok szerint könnyen rendszerezzezzük és csoportosítsuk, most már készen állunk arra, hogy a csoportosított és szűrt projekteket különböző vállalati számításokkal és üzleti forgatókönyvekkel integráljuk.



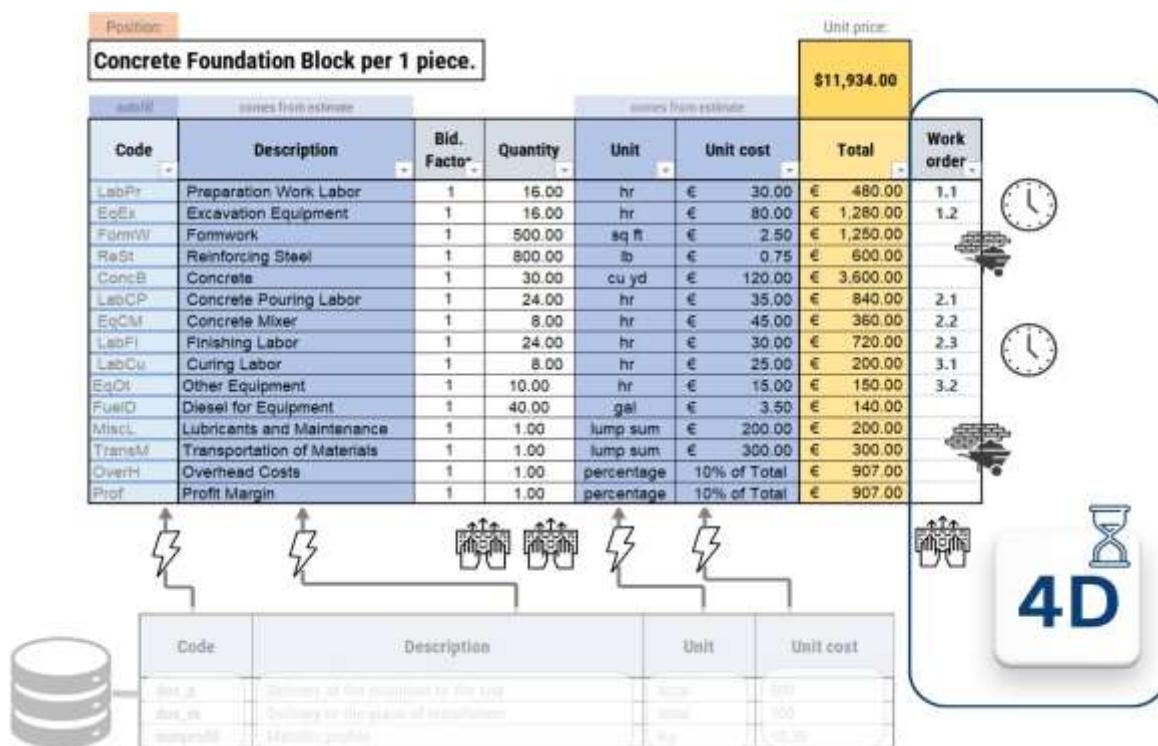
FEJEZET 5.3.

4D, 6D -8D ÉS A SZÉNDIOXID-KIBOCSÁTÁS KISZÁMÍTÁSA CO₂

4D modell: az idő beépítése az építési becslésekbe

A költségszámítás mellett a tervezési adatok egyik legfontosabb alkalmazása az építőiparban az időparaméterek meghatározása - mind az egyes építési műveletekre, mind a teljes projektre vonatkozóan. Az erőforrás-alapú becslési módszert és a hozzá tartozó számítási adatbázist, amelyet az előző, "Számítások és becslések építési projektekhez" című fejezetben részletesen tárgyaltunk, gyakran használják az automatizált időszámítás és a munka elvégzésére vonatkozó ütemterv elkészítésének alapjaként.

Az erőforrás-alapú megközelítés nemcsak az anyagköltségeket, hanem az időforrásokat is figyelembe veszi. A költségszámításban minden folyamathoz hozzárendelhető egy munkamegrendelési attribútum (5.3-1. ábra - Munkamegrendelési paraméter), valamint az adott folyamat végrehajtásához kapcsolódó idő és költség mennyisége. Ezek a paraméterek különösen fontosak az olyan műveletek leírásánál, amelyeknek nincs rögzített piaci ára, és amelyek nem közvetlenül megvásárolhatók - mint például az építőipari berendezések használata, a munkások foglalkoztatása vagy a logisztikai folyamatok (amelyeket általában általában órákban fejeznek ki). Ilyen esetekben a költségeket nem a beszerzési osztály határozza meg, hanem közvetlenül a megvalósító vállalat a belső normák vagy termelési ráták alapján (5.3-1. ábra).



Ábra 5.3-1 A munka kiszámítása az erőforrás-alapú becslési módszerben a munkaórák időzítését is magában foglalja.

Igy a költségszámítási szintű számítások nemcsak az üzemanyag- és anyagköltségeket (beszerzési

költség), hanem a járművezetők, technikusok és segédmunkások által a helyszínen eltöltött időt is tartalmazzák. A megadott példában (5.3-1. ábra) a költségtáblázat egy alapblokk beépítésének költségeit tartalmazza, beleértve a munka alkotó szakaszait, mint például az előkészítés, a keret beépítése és a betonozás, valamint a szükséges anyagokat és munkaerőt. Az egyes műveleteknek, mint például az előkészítő munkálatoknak azonban lehet, hogy nincsenek anyagköltségei, de jelentős, munkaórákban kifejezett ideiglenes munkaköltségeket tartalmazhatnak.

Az építkezésen a munkasorrend (a munka ütemezéséhez) megtervezéséhez a "Munkarend" attribútumot kézzel adjuk hozzá a számítási táblázathoz (5.3-1. ábra). Ezt az attribútumot csak olyan tételek esetében kell egy további oszlopban megadni, amelyek mértékegysége időben (óra, nap) van kifejezve. Ez az attribútum a munkakód, a leírás, a mennyiség, a mértékegység ("Unit" paraméter) és a költségek mellett szerepel. A tevékenységek numerikus sorrendje (a "Munkasorrend" paraméter) lehetővé teszi, hogy megállapítsa a feladatok elvégzésének sorrendjét az építkezésen, és ezt használja az ütemezéshez.

Építési ütemterv és annak automatizálása a költségszámítási adatok alapján

Az építési ütemterv a projekt megvalósítása során végrehajtandó munkaterv és folyamatok vizuális megjelenítése. Részletes erőforrás-számítások alapján készül (5.3-1. ábra), ahol az egyes feladatokat-munkákat az erőforrások költsége mellett idő és sorrend szerint ütemezik.

Szemben az átlagolásos megközelítésekkel, ahol az időbecslések az anyagok vagy berendezések telepítésének tipikus óráin alapulnak, az erőforrás-alapú módszerben a tervezés a költségszámításban szereplő tényleges adatokon alapul. A munkaköltségekkel kapcsolatos becslés minden egyes tétele az alkalmazott naptáron alapul, amely figyelembe veszi az erőforrás-felhasználás tényleges feltételeit a munkaidőszak alatt. A produktív órák kiigazítása a kalkuláció szintjén alkalmazott együtthatókon keresztül (5.3-1 ábra "Bid. Factor" paraméter), lehetővé teszi a termelékenységben mutatkozó különbségek és a munka ütemezését befolyásoló szezonális sajátosságok figyelembevételét.

Az építési ütemterv Gantt-diagramban szereplő folyamat kezdeti és befejező időpontjának meghatározásához az egyes tételek időattribútumának értékeit vesszük az alapblokkok kalkulációjából, és megszorozzuk a blokkok számával (ebben az esetben a beton alapblokkok számával). Ez a számítás adja az egyes feladatok időtartamát. Ezután ezeket az időtartamokat egy idővonalon ábrázoljuk, a projekt kezdeti dátumától kezdve, hogy egy grafikont hozzunk létre, és az eredmény egy vizuális ábrázolás, amely megmutatja, hogy az egyes feladatoknak mikor kell elkezdődniük és befejeződniük. A folyamatok "Munkasorrend" paramétere emellett lehetővé teszi számunkra, hogy megértsük, hogy a munkafolyamat párhuzamosan ("Munkasorrend" pl. 1.1-1.1) vagy szekvenciálisan (1.1-1.2) folyik-e le.

A Gantt-diagram egy grafikus eszköz a tervezéshez és a projektmenedzsmenthez, amely a feladatokat vízszintes sávokként ábrázolja egy idővonalon. minden egyes sáv a feladat időtartamát, kezdetét és végét jelöli.

Az ütemterv vagy Gantt-diagram segít a projektmenedzsereknek és a munkásoknak világosan megérteni, hogy mikor és milyen sorrendben kell elvégezni az építés különböző fázisait, biztosítva az erőforrások hatékony felhasználását és a határidők betartását.

Mutassuk be három beton alapblokk beépítésének ütemtervét a fenti táblázat számításai alapján. A fenti példában szereplő költségtáblázat (5.3-1. ábra) segítségével kérjük meg az LLM-t, hogy ütemezze a 3

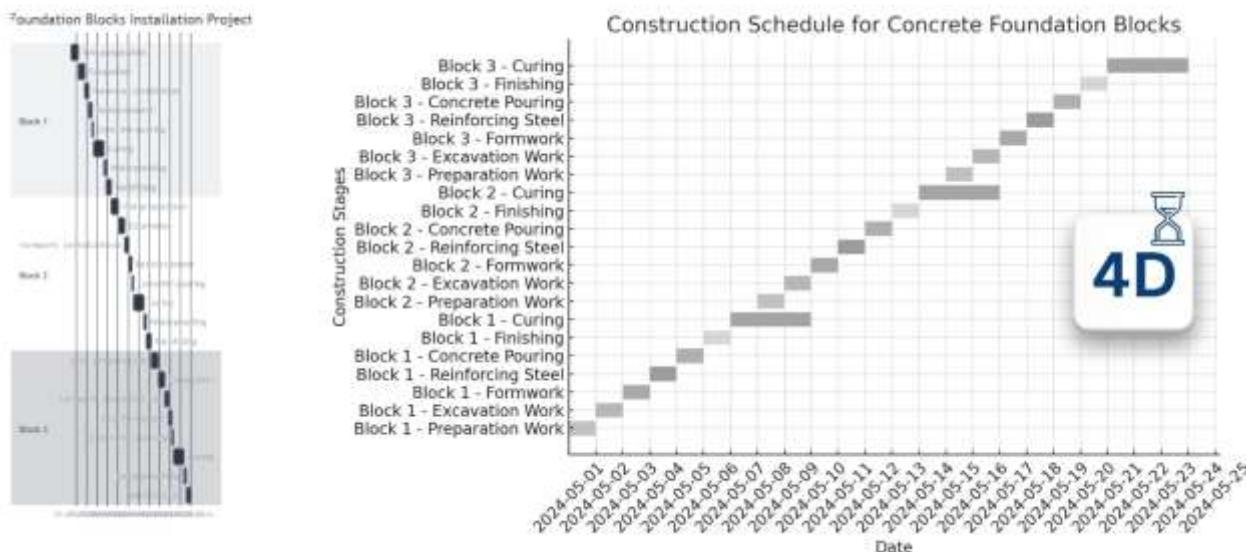
alapblokk elem beépítését, pl. 2024. május elsejére.

A kalkuláció elküldéséhez az LLM, feltölthetjük a kalkulációs táblázatot XLSX formátumban, vagy egyszerűen közvetlenül az LLM csevegésbe beilleszthetjük a kalkuláció JPEG képének képernyőképét (5.3-1. ábra). Az LLM önállóan keres egy könyvtárat a táblázat képének vizualizálásához, és a táblázatban szereplő tevékenységek időattribútumait a mennyiségükkel megszorozva az összes adatot egy grafikonba illeszti.

❷ Küldjön szöveges kérést az LLM címre:

Készítsen Gantt-diagram formájában egy munkaprogramot az időhöz kapcsolódó táblázat értékeinek felhasználásával (5.3-1 ábra JPEG formátumban a függelékben) a 3 alapblokk beépítésére. Az egyes blokkok munkáit egymás után kell elvégezni. A munkálatok kezdetét 2024. május 01-től kell meghatározni. ↴

❸ LLM válasza

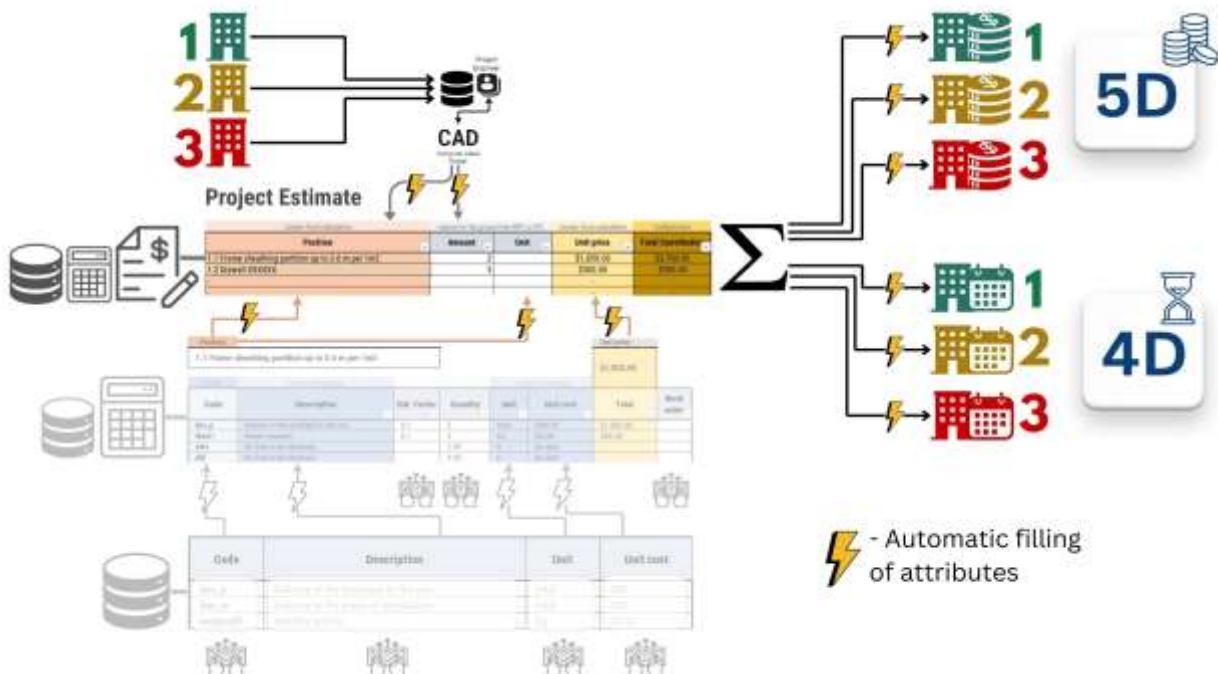


Ábra 5.3-2 Több LLM által automatikusan generált Gantt-diagram mutatja három betontömb építésének szakaszait, a promptból származó feltételek szerint.

Az így kapott grafikon (5.3-2. ábra) egy idődiagram, amelyben minden egyes vízszintes sáv az alaptömbön végzett munka egy bizonyos szakaszának felel meg, és a műveletek sorrendjét mutatja (a "Munkasorrend" paraméter), mint például az előkészítés, a földmunka, a zsaluzat telepítése, a megerősítés, a betonozás és a befejezés, azaz azokat a folyamatokat, amelyek a számításokban kitöltött időparaméterekkel és sorrenddel rendelkeznek.

Egy ilyen ütemterv (5.3-2. ábra) nem veszi figyelembe a munkanapokkal, műszakokkal vagy munkaidőszabványokkal kapcsolatos korlátozásokat, hanem kizárolag a folyamat koncepcionális szemléltetésére szolgál. A munka párhuzamosságát tükröző pontos ütemterv a csevegőszobán belül megfelelő promtakkal vagy további utasításokkal egészíthető ki

Egyetlen költségszámítás (5.3-1. ábra) segítségével, a 3D -geometriából származó térfogati attribútumoknak köszönhetően lehetőség van minden projekt költségek automatikus becslésére az automatizált becslésekben keresztül, minden pedig a csoportok időjellemzőinek táblázatos vagy grafikus formában történő kiszámítására a különböző projektváltozatokra (5.3-3. ábra).



Ábra 5.3-3 Automatikus számítás, lehetővé teszi az azonnali és automatikus költség- és időprognózist a különböző projektlehetőségekre.

A modern moduláris ERP -rendszer (5.4-4. ábra), amelyek CAD modellekből töltik be az adatokat, hasonló automatikus időszámítási módszereket alkalmaznak, amelyek jelentősen csökkentik a döntéshozatali folyamatot. Ez lehetővé teszi a munka ütemezésének azonnali és pontos megtervezését, valamint a projekt megvalósításában az összes feladat elvégzéséhez szükséges teljes idő kiszámítását a valós árak figyelembevételével.

Kibővített attribútumrétegek 6D -8D: az energiahatékonyságtól a biztonság garantálásáig.

A 6D, 7D és 8D az információmodellezés kiterjesztett szintjei, amelyek mindegyike további attribútumrétegekkel járul hozzá az átfogó projektinformációs modellhez, amelynek alapja a 3D -modell attribútumai azok számával és terjedelmével. minden egyes további réteg olyan speciális paraméterekkel járul hozzá, amelyek más rendszerekben, mint például az ingatlangazdálkodási rendszerek (PMS), a számítógéppel támogatott létesítménygazdálkodás (CAFM), az építési projektmenedzsment (CPM) és a biztonságirányítási rendszerek (SMS) további csoportosításához vagy további azonosításához szükségesek.



5.3-4. ábra A adatinformációs modellben a 6D, 7D és 8D attribútumok a projekt különböző szempontjainak figyelembevételét bővítik, az energiahatékonyságtól a biztonságig.

- **A 6D** oldalon az elemek geometriai és térfogati attribútumait tartalmazó projektadatbázis (vagy adatkeret (4.1-13. ábra)) mellett a környezeti fenntarthatóságra vonatkozó információk (attribútum-oszlopok) is szerepelnek. Ez magában foglalja az energiahatékonysággal, a szén-dioxid-kibocsátással, az anyagok újrahasznosításával és a környezetbarát technológiák alkalmazásával kapcsolatos információkat. Ezek az adatok lehetővé teszik a projekt környezeti hatásainak értékelését, a projektdöntések optimalizálását és a fenntartható fejlődési célok elérését.
- **7D** attribútumok kiegészítik az épületek karbantartásának irányításához szükséges attribútumokat. Ezek a karbantartási ütemtervekre, az alkatrészek életciklusára, a műszaki dokumentációra és a javítási előzményekre vonatkozó adatok. Ez az információkészlet biztosítja, hogy a modell integrálható legyen a karbantartási rendszerekkel (CAFM, AMS), lehetővé teszi a karbantartás hatékony tervezését, a berendezések cseréjét, és támogatást nyújt a létesítmény teljes életciklusa során.
- **8D** további attribútumréteg, - a biztonsággal kapcsolatos információkat tartalmaz - mind az építési szakaszban, mind a későbbi üzemeltetés során. A modell tartalmazza a személyi biztonsági intézkedéseket, a vészhelyzeti utasításokat, az evakuációs rendszereket és a tűzvédelmi követelményeket. Ezeknek az adatoknak a digitális modellbe való integrálása segít a kockázatok előzetes figyelembevételében, valamint az egészségvédelmi és biztonsági követelményeket figyelembe vevő építészeti, mérnöki és szervezési megoldások kidolgozásában.

Strukturált táblázatos formában a 4D - 8D rétegek további attribútumokat jelentenek oszlopok formájában, kitöltött értékekkel (5.3-5. ábra), amelyek a már kitöltött 3D -modell attribútumokhoz, mint például a név, kategória, típus és térfogati jellemzők, adódnak hozzá. A 6D, 7D és 8D attribútumrétegek értékei további szöveges és numerikus adatokat tartalmaznak, mint például az újrahasznosítási százalék, a szénlábnyom, a garanciaidő, a csereciklus, a telepítési dátum, a biztonsági protokollok stb.



ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO ₂ e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO ₂ e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO ₂ e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO ₂ e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

A 5.3-5. ábra 6D -8D attribútumrétegeket ad hozzá az adatinformációs modellhez, amely már tartalmazza a 3D -modell geometriai és térfogati attribútumait.

Az új ablakunk (4.4-1. ábra) esetében a W-NEW azonosítóval rendelkező elem (5.3-5. ábra) a következő 3D -8D attribútumokkal rendelkezhet:

3D -attributes - CAD rendszerekből nyert geometriai információk:

- "Típusnév" - "Window" elem
- "Szélesség" - 120 cm
- Ezen kívül egy elem "Bounding Box" pontjait vagy a "geometria BREP / MESH" külön attribútumként adhatja meg.

A 6D jellemzői - környezeti fenntarthatóság:

- 90 százalékos újrahasznosítási arány
- "Karbonlábnyom - 1,622 kg CO₂

Attribútumok 7D - objektumkezelési adatok:

- "Jótállási idő" - 8 év
- A "Replacement Cycle" 20 éves
- "Karbantartás" - évente szükséges

A 8D jellemzői - az épületek biztonságos használatának és üzemeltetésének biztosítása:

- "Telepített" ablak - az "XYZ Windows" vállalat által
- "Biztonsági szabvány" - megfelel az ISO 45001 szabványnak

Az adatbázisban vagy adatkészletben (5.3-5. ábra) rögzített összes paraméterre a különböző részlegek szakembereinek csoportosításhoz, kereséshez vagy számításokhoz van szükségük. A projektobjektumok ilyen többdimenziós, attribútumalapú leírása teljes képet nyújt az életciklusukról, az üzemeltetési követelményekről és sok más, a projekttervezéshez, -építéshez és -üzemeltetéshez szükséges szempontról.

A CO₂ becslése és az építési projektek szén-dioxid-kibocsátásának kiszámítása

Az építési projektek fenntarthatóságának téma mellett a 6D szakaszban (5.3-5. ábra) a modern építőipar a projektek környezeti fenntarthatóságára összpontosít, ahol az egyik legfontosabb szempont a projekt életciklusának szakaszaiban (pl. gyártás és telepítés) keletkező szén-dioxid CO₂ kibocsátás felmérése és minimalizálása lesz.

Az építőanyagok szén-dioxid-kibocsátásának becslése és kiszámítása olyan folyamat, amelynek során a teljes szén-dioxid-kibocsátást úgy határozzák meg, hogy a projektben felhasznált elemek vagy elemcsoportok térfogati jellemzőit megsorozzák a kategóriának megfelelő szén-dioxid-kibocsátási tényezővel.

A szén-dioxid-kibocsátás figyelembevétele az építési projektek értékelésénél a szélesebb körű ESG (környezetvédelmi, társadalmi és kormányzási) kritériumok részeként új komplexitási szintet ad az elemzéshez. Ez különösen fontos az ügyfél-befektető számára a vonatkozó tanúsítványok megszerzésében, mint például a LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) vagy DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). E tanúsítványok valamelyikének megszerzése jelentősen növelte az ingatlan piacképességét, egyszerűsítheti az üzembe helyezést, és biztosíthatja a fenntarthatósági szempontú bérzők (ESG) követelményeinek való megfelelést. A projekt követelményeitől függően a HQE (Haute Qualité Environnementale, a francia zöldépítési szabvány), a WELL (WELL Building Standard, a felhasználók egészségére és kényelmére összpontosít) és a GRESB (Global Real Estate Sustainability Benchmark) is alkalmazható

Környezeti, társadalmi és kormányzási **ESG** (környezeti, társadalmi és kormányzási) elvek széles köre, amelyek segítségével értékelhető egy vállalkozás vállalatirányítási, társadalmi és környezeti hatása mind belső, mind külső szinten.

Az ESG, amelyet eredetileg a 2000-es évek elején a pénzügyi alapok fejlesztettek ki, hogy a befektetők számára információkat nyújtsanak az átfogó környezeti, társadalmi és irányítási kritériumokról, mára a vállalatok és a projektek - köztük az építési projektek - értékelésének kulcsfontosságú mutatójává vált. A nagy tanácsadó cégek kutatásai szerint a környezeti, társadalmi és kormányzási (ESG) szempontok az építőipar szerves részévé válnak.

Az EY (2023) "The Path to Carbon Neutrality" (Út a karbonsemlegességhoz) című tanulmánya szerint az ESG -elveket aktívan megvalósító vállalatok nemcsak a hosszú távú kockázatokat csökkentik, hanem üzleti modelljeik hatékonyságát is növelik, ami különösen fontos a piacok globális átalakulásában [103]. A PwC ESG-tudatosságról szóló jelentése megállapítja, hogy a vállalatok tudatában vannak az ESG -faktorok fontosságának 67% és 97% között mozog, és a legtöbb szervezet úgy látja, hogy ezek a tendenciák kulcsfontosságúak a jövőbeli fenntarthatóság szempontjából [104], és a vállalatok többsége jelentős nyomást gyakorol az érdekkelt felek részéről az ESG-elvek integrálására.

Így az ESG -elvek beépítése az építési projektekbe nemcsak a nemzetközi fenntarthatósági tanúsítványok - mint például a LEED, BREEAM, DGNB - megszerzéséhez járul hozzá, hanem biztosítja az ágazatban működő vállalatok hosszú távú fenntarthatóságát és versenyképességét is.

Az építési projekt teljes szénlábjnyomát befolyásoló egyik legjelentősebb tényező az építőanyagok és alkatrészek gyártási és logisztikai szakasza. A helyszínen felhasznált anyagok gyakran döntő hatással vannak a teljes CO₂-kibocsátásra, különösen a projekt életciklusának korai szakaszában - a nyersanyagok kitermelésétől az építkezés helyszínére történő szállításig.

A kibocsátások kategóriánként vagy épületelemtípusonként történő kiszámításához olyan szén-dioxid-kibocsátási referencia-tényezőket kell alkalmazni, amelyek tükrözik a különböző anyagok előállítása során keletkező CO₂ mennyiséget. Ilyen anyagok például a beton, a téglák, az újrahasznosított acél, az alumínium és mások. Ezeket az értékeket általában neves forrásokból és nemzetközi adatbázisokból nyerik ki, mint például a UK ICE 2015 (Inventory of Carbon and Energy) és az US EPA 2006 (U.S. Environmental Protection Agency) [105]. A következő táblázat (5.3-6. ábra) összefoglalja az alapszintű kibocsátási tényezőket egy sor gyakori építőanyagra vonatkozóan. minden anyaghoz két kulcsfontosságú paramétert adunk meg: a fajlagos CO₂-kibocsátást (kilogrammonként gramm anyagra vetítve) és a térfogat-súly átváltási tényezőket (kilogrammonként köbméterenként), amelyek szükségesek a számítások tervezési modellbe való integrálásához és a QTO adatcsoportosítással való összekapcsolásához.



Carbon Emitted in Production		UK ICE Database [2015] USEPA [2006]	UK ICE Database [2015] USEPA [2006]	Coefficient m3 to kg
Material	Abbreviated	Process Emissions (kg CO2e/kg of product) (K1)	Process Emissions (kg CO2e/kg of product) (K2)	Kg / m3 (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

5.3-6. ábra A különböző építőanyagok előállítása során kibocsátott szén-dioxid mennyisége az Egyesült Királyság ICE és az Egyesült Államok EPA adatbázisa szerint.

A projekt teljes CO₂ kibocsátásának kiszámításához - a 4D és az 5D számításokhoz hasonlóan - meg kell határozni az egyes létesítménycsoportok attribútumainak mennyiséget. Ez mennyiségi elemzési eszközökkel (QTO) végezhető el az attribútum-mennyiségek köbméterben történő meghatározásával, amint azt a mennyiségi átvételeiről szóló fejezetben részletesen tárgyaljuk. Ezeket a térfogatokat ezután meg kell szorozni az egyes anyagcsoportok "CO₂-folyamatok kibocsátása" attribútumának megfelelő együtthatóival.

- ❸ Vegyük ki automatikusan a CAD (BIM) projektből az elemtípusok szerinti térfogattáblázatot a projekt összes adatának csoportosításával, ahogyan azt az előző fejezetekben már megtettük.

A feladat elvégzéséhez hivatkozzunk az LLM.

Kérjük, csoportosítsa a DataFrame táblázatot a CAD (BIM) projektből a "Tárgynév" (vagy "Típus") oszlopparaméter szerint, és mutassa meg az egyes csoportokban lévő elemek számát, valamint foglalja össze a "Térfogat" paramétert a típus összes elemére vonatkozóan. ↴

■ LLM válasza :

The screenshot shows a Jupyter Notebook environment. On the left, there's a sidebar with icons for 'Create in LLM Chat' (with a brain icon), 'Data' (with a bar chart icon), 'f(x)' (with a function icon), and 'Output' (with a table icon). Below these are buttons for 'Run in IDE' (with a PC icon) and 'Run' (with a play button icon). The main area contains a code cell and its output. The code imports pandas, reads a CSV file, groups by 'ObjectType', and calculates the sum of 'Volume'. The output is a grouped DataFrame:

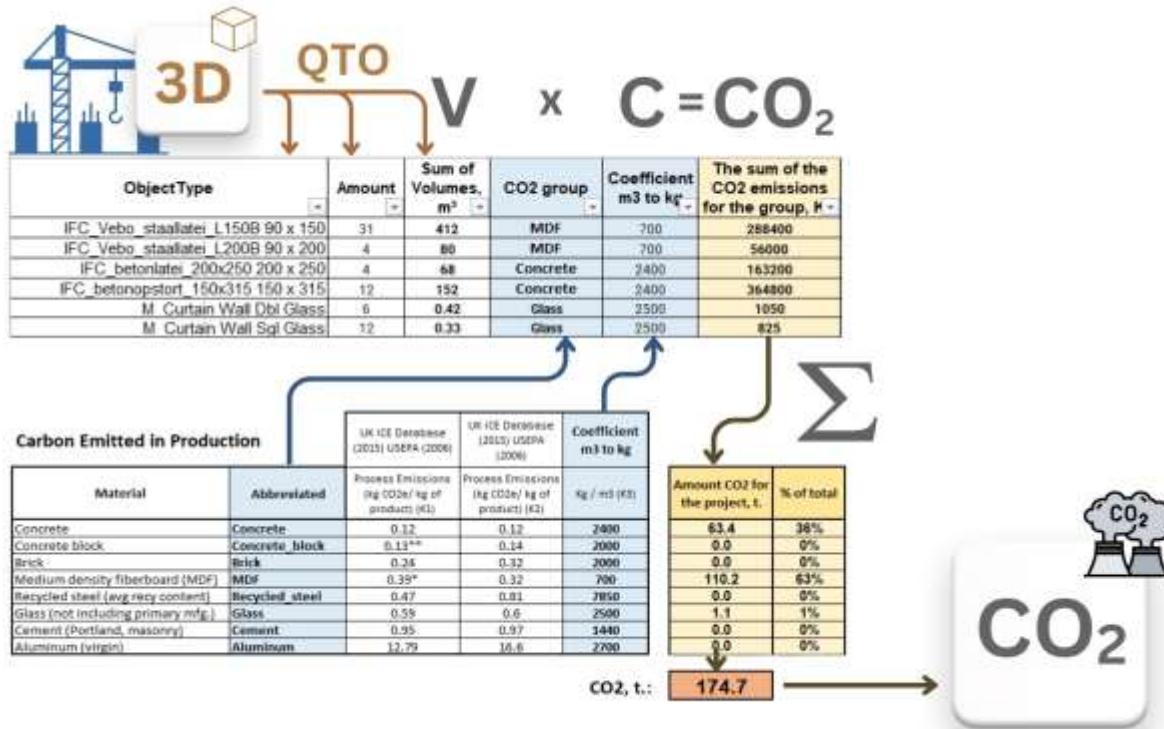
ObjectType	Amount	Sum of Volumes, m ³
IFC_Vebo_stallatei_L150B_90 x 150	31	412
IFC_Vebo_stallatei_L200B_90 x 200	4	80
IFC_betonlatei_200x250_200 x 250	4	68
IFC_betonopstort_150x315_150 x 315	12	152
M_Curtain_Wall_Dbi_Glass	6	0.42
M_Curtain_Wall_Sgl_Glass	12	0.33

5.3-7 ábra Az LLM generált kódja típusok (ObjectType) szerint csoportosította számunkra a projekt entitásait a "Volume" attribútummal.

A teljes projekt teljes CO₂ kibocsátásának automatizált kiszámításához egyszerűen állítsa be az automatikus adattérképezést a táblázatban, vagy kapcsolja össze manuálisan az elemtípusokat (5.3-7. ábra) a megfelelő anyagtípusokkal (5.3-6. ábra) a kibocsátási tényező táblázatból. A kész táblázat a kibocsátási tényezőket és képleteket, valamint a CAD-formátumokból (BIM) a térfogatok lekérdezéséhez és a CO₂-meghatározás automatizálásához szükséges kód megtalálható a GitHubon a "CO₂_calculating-the-embodyied-carbon" kereséssel. DataDrivenConstruction." [106].

Így a CAD-adatbázisból származó QTO-elemek csoportosítását követő adatintegráció lehetővé teszi a szén-dioxid-kibocsátás automatikus kiszámítását (5.3-8. ábra) különböző tervezési lehetőségek esetén. Ez lehetővé teszi a különböző változatokban alkalmazott különböző anyagok hatásának elemzését, és csak olyan megoldások kiválasztását, amelyek megfelelnek megrendelő CO₂-kibocsátási követelményeinek, hogy az épület üzembe helyezésekor egy adott tanúsítványt kapjon

A CO₂-kibocsátás becslése a tényezőknek a csoportosított projektelemek mennyiségével való megszorzásával tipikus példája annak a feladatnak, amelyet egy építőipari vállalat végez, amikor egy létesítmény számára ESG-minősítést szerez (pl. LEED-tanúsítvány).

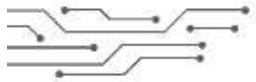


A 5.3-8. ábra A QTO-csoportok integrálása a CAD-adatbázisokból biztosítja a pontosságot és az automatizációt a végső CO₂-kibocsátás becslésének levezetésében.

Hasonlóképpen, az elemcsoportok mennyiségének meghatározásával számításokat végezhetünk az anyagellenőrzés és logisztika, a minőségellenőrzés és -irányítás, az energiamodellezés és -elemzés, valamint egy sor más feladat elvégzésére, hogy új attribútumállapotot (paramétert kapunk a táblázatban) minden elemcsoportra, minden teljes projektre vonatkozóan.

Ha a vállalaton belül az ilyen számítási folyamatok száma nőni kezd, felmerül a kérdés, hogy szükség van-e az ilyen számítások automatizálására és a számítások eredményeinek a vállalat folyamataiba és adatkezelési rendszereibe való beépítésére.

A teljes megoldás összetettsége miatt az építőipari közép- és nagyvállalatok az ilyen automatizálást kiszervezik az ERP (vagy PMIS) rendszerfejlesztő cégeknek. A fejlesztő cégek egyetlen átfogó, moduláris rendszert hoznak létre a nagy ügyfelek számára, amely számos különböző információs réteget kezel, beleértve az anyag- és erőforrás-számításokat is.



FEJEZET 5.4.

ÉPÍTŐIPARI ERP ÉS PMIS RENDSZEREK

Építőipari ERP -rendszerek a számítások és becslések példáján

A moduláris ERP rendszerek egyetlen átfogó rendszerbe integrálják a különböző attribútum (információs) rétegeket és adatáramlásokat, lehetővé téve a projektmenedzserek számára, hogy egyetlen platformon belül, szinkronizált módon kezeljék az erőforrásokat, a pénzügyeket, a logisztikát és a projekt egyéb aspektusait. Az építőipari ERP-rendszer az építési projektek "agyaként" működik, automatizálással egyszerűsíti az ismétlődő folyamatokat, átláthatóságot és ellenőrzést biztosít az építési folyamat során.

Az építőipari ERP -rendszerek (Enterprise Resource Planning) olyan átfogó szoftvermegoldások, amelyeket az építési folyamat különböző aspektusainak kezelésére és optimalizálására terveztek. Az építőipari ERP-rendszerök középpontjában a költségszámítás és az ütemezés kezelésére szolgáló modulok állnak, így a hatékony erőforrás-tervezés fontos eszközeivé válnak.

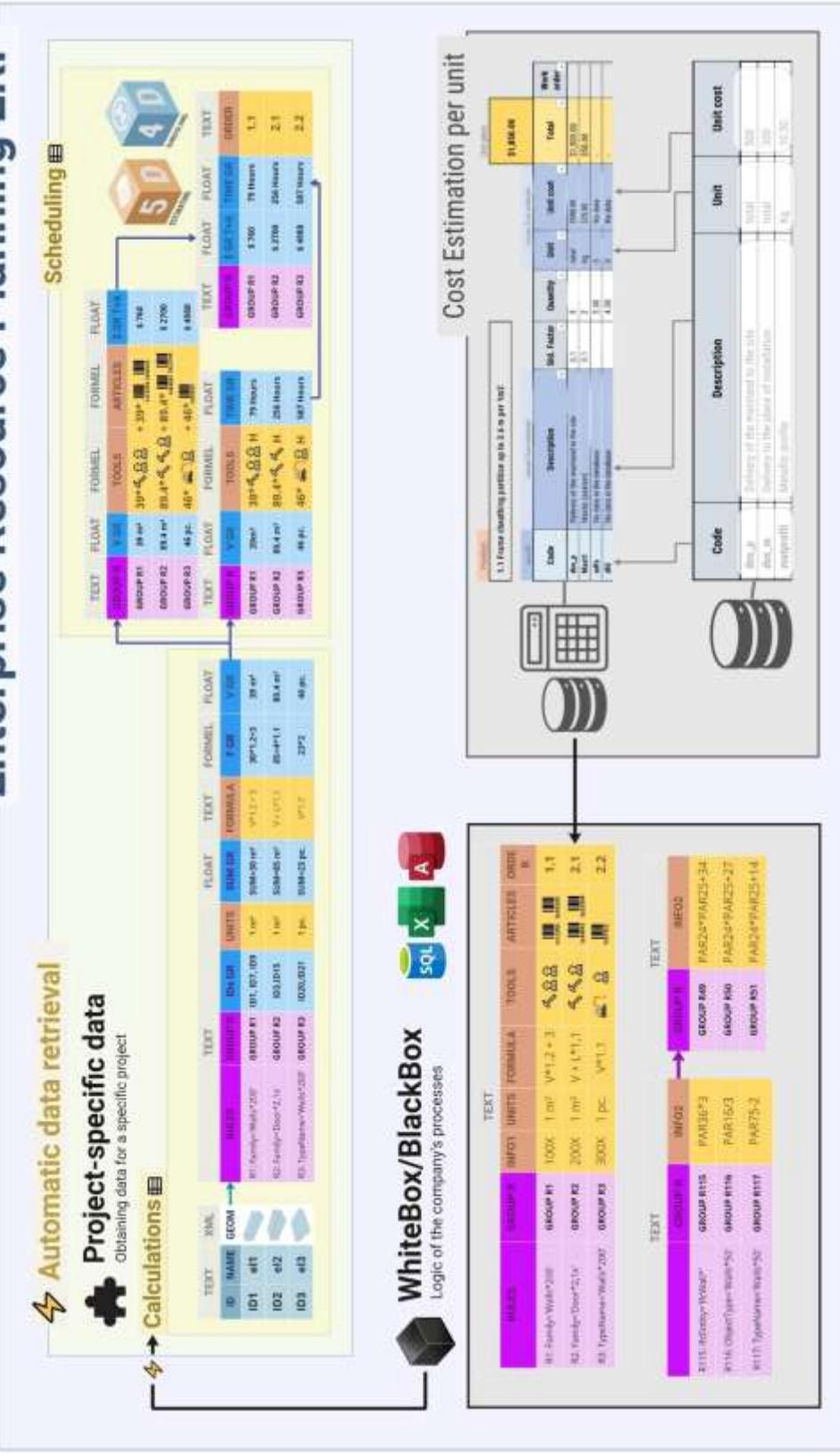
Az ERP modulok -rendszerek lehetővé teszik a felhasználók számára, hogy strukturált módon adjanak meg, dolgozzanak fel és elemezzék adatokat a projekt különböző aspektusaira vonatkozóan, amelyek magukban foglalhatják az anyag- és bérköltségek elszámolását, a berendezések felhasználását, a logisztikai irányítást, a humán erőforrásokat, a kapcsolatokat és egyéb építési tevékenységeket.

A rendszer egyik funkcionális blokkja az üzleti logika automatizálásának modulja - BlackBox/WhiteBox, amely a folyamatirányító központ szerepét tölti be.

A BlackBox /WhiteBox lehetővé teszi az ERP -rendszer használó szakemberek számára, hogy a hozzáférési jogok révén rugalmasan kezeljék az üzlet különböző aspektusait, amelyeket más felhasználók vagy rendszergazdák már előre konfiguráltak. Az ERP-rendszerrel összefüggésben a *BlackBox* és *WhiteBox* kifejezések a rendszer belső logikájának átláthatósági és ellenőrizhetőségi szintjére utalnak:

- **BlackBox ("fekete doboz")** - a felhasználó a felületen keresztül lép kapcsolatba a rendszerrel, anélkül, hogy hozzáérne a folyamatvégrehajtás belső logikájához. A rendszer a végfelhasználó elől elrejtett, előre meghatározott szabályok alapján saját maga végzi el a számításokat. Beadja az adatokat és megkapja az eredményt anélkül, hogy tudná, milyen attribútumokat vagy együtthatókat használtak belül.
- **WhiteBox ("fehér doboz")** - a folyamatlogika megtekinthető, testreszabható és módosítható. A haladó felhasználók, rendszergazdák vagy integrátorok kézzel határozhatsák meg az adatfeldolgozási algoritmusokat, számítási szabályokat és a projektek közötti interakciós forgatókönyveket.

Enterprise Resource Planning ERP



Ábra 5.4-1 Építőipari ERP -rendszer architektúrája, hogy becsléseket és ütemtervezeket kapunk a munka mennyiségeinek attribútumainak kézi kitöltésével.

Egy példa erre az, amikor egy tapasztalt felhasználó vagy rendszergazda szabályt állít be: egy becslésben mely attribútumokat kell egymás között megszorozni vagy egy bizonyos jellemző alapján csoportosítani, és a végeredményt hol kell rögzíteni. Később a kevésbé képzett szakemberek, például a becsléssel foglalkozó mérnökök egyszerűen betöltenek új adatokat az ERP-be a felhasználói felületen keresztül - és kész becsléseket, ütemterveket vagy specifikációkat kapnak anélkül, hogy kódot kellene írniuk vagy érteniük kellene a logika technikai részleteit.

Az előző fejezetekben a számítási és logikai modulokat az LLM kölcsönhatás összefüggésében tárgyaltuk. Egy ERP-környezetben az ilyen számítások és átalakítások a gombokból és űrlapokból álló felület mögé rejtett modulokon belül zajlanak.

A következő példában (5.4-1. ábra) az ERP -rendszer rendszergazdája a BlackBox /WhiteBox modulban szabályokat definiált a becslésekkel származó entitások attribútumainak a QTO csoportosítás attribútumaival való egyeztetésére. Ennek a (vezető vagy rendszergazda által konfigurált) BlackBox/WhiteBox modulnak köszönhetően a felhasználó (becslő vagy mérnök) az ERP felhasználói felületen keresztül manuálisan hozzáadott mennyiségi vagy mennyiségi attribútummal automatikusan megkapja a kész becsléseket és munkaterveket. Ily módon az előző fejezetekben tárgyalt számítási és becslésgenerálási folyamatok kód segítségével, az ERP-n belül, félautomatizált futószalaggá válnak.

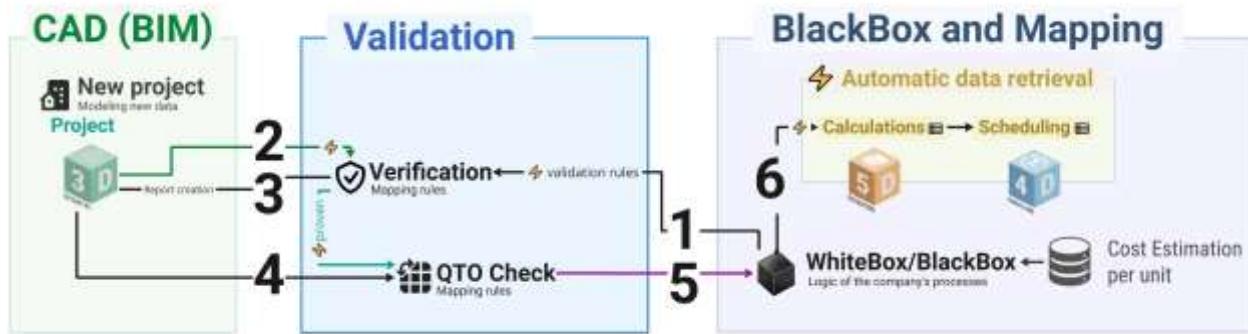
Ennek a félautomatizált folyamatnak a CAD (BIM) modellekből származó volumetrikus attribútumokhoz való csatlakoztatása (4.1-13. ábra), például a CAD-projektnek az ERP modulba való betöltése révén, amely erre a célra előre be van állítva, az adatáramlást olyan szinkronizált mechanizmussá alakítja, amely képes önállóan és azonnal frissíteni az egyes elemcsoportok vagy a teljes projekt értékét a tervezési fázisban bekövetkező változásokra reagálva, amikor a CAD-modellt betöltik az ERP-be.

A CAD (BIM) és az ERP rendszerek közötti automatizált adatáramlás (5.4-2. ábra) létrehozása érdekében a CAD (BIM) modelladatbázisokból származó adatokra vonatkozó alapvető folyamatokat és követelményeket strukturált módon kell meghatározni, ahogyan azt már a fenti "Követelmények és adatminőségbiztosítás " című fejezetben tárgyaltuk. Ez a folyamat az ERP-ben hasonló lépésekre oszlik:

- **Érvényesítési szabályok létrehozása (1)**, amelyek fontos szerepet játszanak az ERP -rendszerbe kerülő adatok pontosságának biztosításában. Az érvényesítési szabályok szűrőként szolgálnak, amelyek az entitások és attribútumaik érvényesítésére szolgálnak, és csak a követelményeknek megfelelő elemek kerülhetnek be a rendszerbe. Az ellenőrzésről és érvényesítésről többet megtudhat a "A követelmények létrehozása és a adatminőség érvényesítése " című fejezetben.
- Ezután az ERP oldalon belül **egy ellenőrzési folyamat (2) zajlik**, amely megerősíti, hogy a projektelemek attribútumaikkal és értékeikkel együtt helyesen lettek létrehozva, és készen állnak a következő feldolgozási lépésekre.
- Ha a hiányos attribútumadatokkal problémák merülnek fel, **akkor jelentés (3) készül**, és a projektet a javításra vonatkozó utasításokkal együtt elküldi átdolgozásra, amíg készen nem áll a következő iterációra.
- Miután a projektadatokat validálták és ellenőrizték, azokat egy másik ERP-modulban(4) modulban felhasználják a **mennyiségi átvételei táblázatok (QTO) létrehozására**, amelyek az előzetesen generált szabályoknak (WhiteBox/BlackBox) megfelelően mennyiségi attribútumokat hoznak létre az entitáscsoportokhoz, anyagokhoz és erőforrásokhoz.
- Az egyeztetési szabályok vagy a QTO alapján csoportosított adatok automatikusan **integrálódnak**

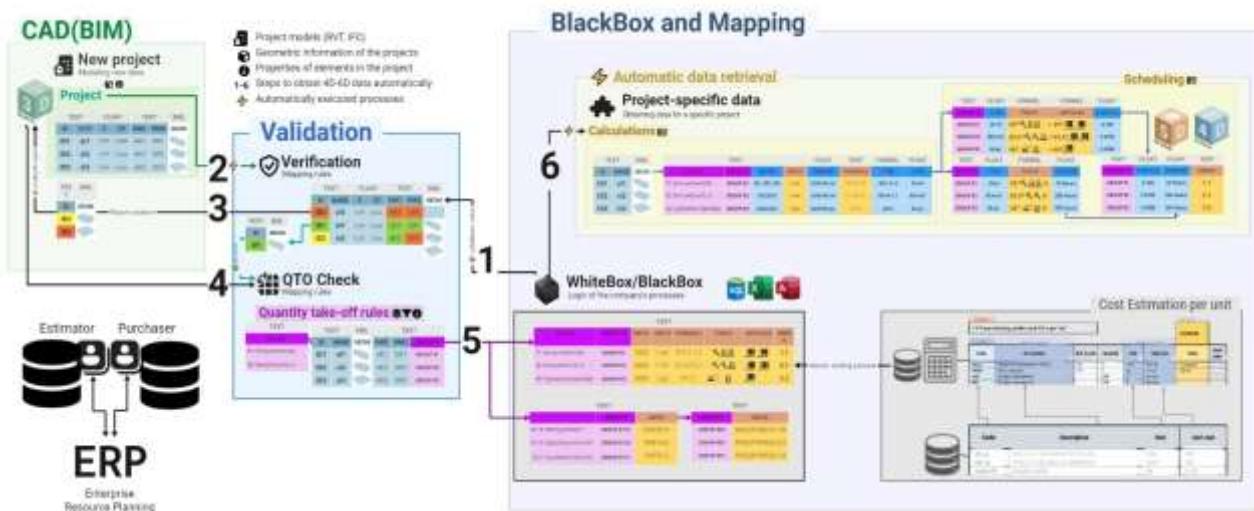
a számításokkal (pl. költség és idő) (5).

- Az ERP -rendszer utolsó lépésében a felhasználó a QTO-táblából származó hatókör-attribútumok és a folyamat-táblák (pl. becsült tételek) attribútumainak megszorzásával **automatikusan számítási eredményeket (6)** (pl. költségbecsléseket, munkarendeket vagy CO₂-kibocsátásokat) **generál** az egyes entitáscsoportokra és a projekt egészére vonatkozóan.

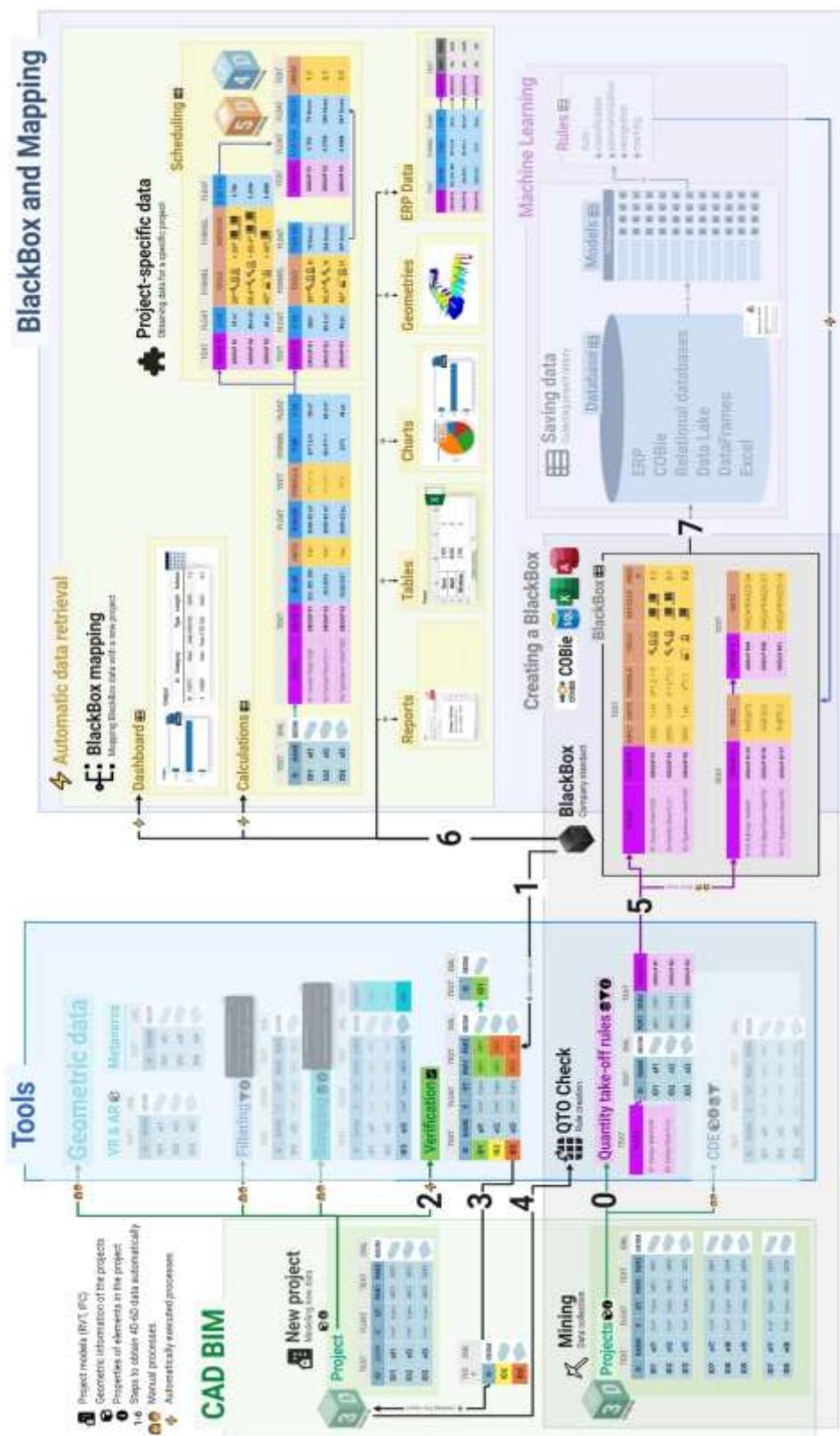


Ábra 5.4-2 Egy építőipari ERP -rendszer felépítése CAD (BIM) segítségével, az érvényesítési szabályok létrehozásától (1) a költségek és a munkaprogramok automatikus kiszámításáig (5-6).

A moduláris ERP -rendszerben a folyamatok integrálása egy felhasználói felületet tartalmazó szoftver segítségével történik. A felület mögött található a back-end, ahol strukturált táblázatok dolgozzák fel az adatokat a vezető vagy a rendszergazda által előre beállított különböző műveletek elvégzésével. Ennek eredményeképpen a felhasználó az előre meghatározott és testre szabott automatizálási logikának köszönhetően (a BlackBox /WhiteBox modulokban) félautomatikusan előkészített dokumentumokat kap, amelyek megfelelnek a feladatainak



Ábra 5.4-3 ERP -rendszer segíti a vezetőket és a felhasználókat a szakterületek táblái közötti mozgásban, hogy új adatokat hozzanak létre.



Ábra 5.4-4 ERP -rendszer elemző eszközökkel integrált, és automatizálja a vállalat döntéshozatali folyamatát.

Hasonlóképpen, az ERP -rendszerek folyamatai a kezdetektől a végső számításig (1-6. lépés 5.4-3. ábra) egymással összefüggő lépések láncolatát alkotják, amelyek végső soron átláthatóságot, hatékonyságot és pontosságot biztosítanak a tervezésben.

A modern építőipari ERP -rendszerek nemcsak költség- és ütemterv-számítási modulokat tartalmaznak, hanem több tucat egyéb, előre konfigurált modult is, amelyek általában a dokumentumkezelésre, a projekt előrehaladásának nyomon követésére, a szerződéskezelésre, az ellátási láncre és a logisztikára, valamint más üzleti rendszerekkel és platformokkal való integrációra terjednek ki. Az ERP integrált elemzési eszközei lehetővé teszik a felhasználók számára, hogy automatizálják a projekt KPI-k (KPI - kulcsfontosságú teljesítménymutatók) nyomon követésére szolgáló műszerfalak létrehozását. Ez az építési projekt minden szempontjának központosított és következetes irányítását biztosítja, és számos alkalmazást és rendszert igyekszik egyetlen platformon egyesíteni.

A jövőben az ERP -analitikát a gépi tanulással kombinálva fogják használni a pontosság növelése és a jövőbeli projektjellemzők kiszámításának optimalizálása érdekében. Az ERP-rendszerekből a Big Data (5.4-4. ábra) keretében elemzett és összegyűjtött adatok és attribútumok a jövőben olyan előrejelző modellek létrehozásának alapját képezik majd, amelyek pontosan előre jelzik a lehetséges késedelmeket, kockázatokat vagy például az anyagköltségek lehetséges változásait.

Az ERP alternatívájaként az építőipar gyakran használja a PMIS (Project Management Information System) projektirányítási rendszert, amelyet a feladatok részletes ellenőrzésére terveztek az egyes építési projektek szintjén.

PMIS: Közvetítő az ERP és az építkezés között.

Az ERP -tól eltérően, amely a vállalat üzleti folyamatainak teljes láncolatát lefedi, a PMIS egy adott projekt irányítására, az ütemezés, a költségvetés, az erőforrások és a dokumentáció nyomon követésére összpontosít.

A PMIS (Project Management Information System) egy építési projektmenedzsment szoftver, amelyet a projekt minden aspektusának megtervezésére, nyomon követésére, elemzésére és jelentésére terveztek.

A PMIS lehetővé teszi a dokumentumok, ütemtervek és költségvetések kezelését, és első pillantásra a PMIS az ERP duplikált megoldásának tűnhet, de a legfontosabb különbség az irányítási szint:

- **Az ERP** a vállalat egészének üzleti folyamataira összpontosít: költség-, szerződés-, beszerzési, humánerőforrás- és erőforrás-menedzsment vállalati szinten.
- **A PMIS** az egyedi projektmenedzsmentre összpontosít, részletes tervezést, változásellenőrzést, jelentéstételt és a résztvevők koordinálását biztosítva.

Sok esetben az ERP -rendszerek már elegendő funkcionálitással rendelkeznek, és a PMIS bevezetése inkább kényelmi és vállalati preferencia kérdése. Sok vállalkozó és ügyfél nem azért használja a PMIS-t, mert szükség van rá, hanem mert a szállító vagy egy nagy ügyfél, aki az adatokat egy adott platformon akarja összesíteni, ezt írja elő.

Meg kell említeni, hogy az építési projektmenedzsment nemzetközi terminológiájában más, különálló, népszerű fogalmak is léteznek, mint például a PLM (Product Lifecycle Management) és az EPC és EPC-M

(Engineering, Procurement and Construction Management) - szerződéskötési módszerek az építőiparban.

Ha egy vállalat már használja az ERP projektmenedzsment modulokat, a PMIS bevezetése felesleges, a funkciókat megkettőző kapcsolat lehet. Ha azonban a folyamatok nem automatizáltak és az adatok töredézettek, a PMIS kényelmesebb és könnyebben karbantartható eszköz lehet.

Spekuláció, profit, elszigeteltség és a hiányos átláthatóság az ERP és a PMIS rendszerekben

Az interfészek és eljárások külső egyszerűsége ellenére az építőipari ERP és PMIS -rendszerek a legtöbb esetben zárt és rugalmatlan megoldások. Az ilyen rendszereket általában egyetlen szállítótól származó, előre konfigurált szoftvercsomagként szállítják, korlátozott hozzáféréssel a belső adatbázisokhoz és folyamatlogikához.

A CAD-(BIM-) szállítók egyre inkább átveszik az ilyen rendszerek fejlesztését és irányítását, mivel adatbázisuk tartalmazza az ERP-rendszerek által igényelt információkat: a projektelemek mennyiségi és térfogati jellemzőit. Ahelyett azonban, hogy nyílt vagy gépileg olvasható formátumban biztosítanának hozzáférést ezekhez az adatokhoz, a szállítók csak korlátozott felhasználói forgatókönyveket és zárt feldolgozási logikát kínálnak - a BlackBox modulokban előre definiálva. Ez csökkenti a rendszer rugalmasságát, és megakadályozza, hogy az egyedi projektfeltételekhez igazítsák.

A korlátozott adatátláthatóság továbbra is az építőipari digitális folyamatok egyik fő kihívása. A zárt adatbázis-architektúra, az építőelemek teljes attribútumkészletéhez való hozzáférés hiánya, a BlackBox automatizálási modulokra való összpontosítás és a nyílt interfészek hiánya jelentősen növeli a dokumentációs bürokrácia kockázatát. Ezek a korlátozások szűk keresztmetszeteket hoznak létre a döntéshozatali folyamatban, megnehezítik az információk ellenőrzését, és utat nyitnak az adatok elrejtésének vagy spekulációjának az ERP/PMIS rendszereken belül. A felhasználók jellemzően csak korlátozott hozzáférést kapnak - legyen az egy lecsupaszított felület vagy egy részleges API - anélkül, hogy közvetlenül kapcsolatba léphetnének az elsődleges adatforrásokkal. Ez különösen kritikus, amikor a CAD-projektekből automatikusan generált paraméterekről van szó, mint például a QTO-számításokhoz használt térfogatok, területek és mennyiségek.

Ennek következtében ahelyett, hogy a folyamatok automatizálásával, a nyílt adatokkal, a tranzakciós költségek csökkentésével és új üzleti modellek létrehozásával törekednének a hatékonyságra, sok építőipari vállalat a külső paraméterek kezelésére összpontosít - a projektköltségeket befolyásoló tényezők, korrekciós tényezők és számítási módszerek manipulálására a zárt ERP/PMIS platformokon. Ez teret ad a spekulációjának, eltorzítja a valós termelési költségeket és csökkenti az építési folyamat valamennyi résztvevője közötti bizalmat

Az építőiparban a nyereség a befejezett projektből származó bevétel és a változó költségek - amelyek magukban foglalják a tervezési, anyag-, munka- és egyéb, a projekt megvalósításához közvetlenül kapcsolódó közvetlen költségeket - közötti különbség. E költségek értékét azonban nemcsak a technológia vagy a logisztika, hanem a számítások gyorsasága és pontossága, valamint a vállalaton belüli vezetői döntések minősége is kulcsfontosságú tényező.

A problémát súlyosbítja, hogy a legtöbb építőipari vállalatnál a költségszámítási folyamatok nem csak az ügyfelek, hanem maguk az alkalmazottak számára is átláthatatlanok maradnak, akik nem tartoznak a becslési vagy pénzügyi osztályokhoz. Ez a zártság kedvez annak, hogy a vállalaton belül kialakuljon a szakemberek egy kiváltságos csoportja - a "pénzügyi szakértelem" hordozói, akik kizárolagos joggal rendelkeznek az ERP/PMIS -rendszerök attribútumainak és korrekciós tényezőinek szerkesztésére. Ezek az alkalmazottak a vállalatvezetőkkel együtt ténylegesen irányíthatják a projekt pénzügyi logikáját.

Ilyen körülmények között a becslők "pénzügyi zsonglörökké" válnak, akik egyensúlyoznak a vállalat nyereségének maximalizálása és az ügyfél számára versenyképes ár fenntartásának szükségessége között. Ugyanakkor el kell kerülniük a kirívó és durva manipulációkat, hogy ne ássák alá a vállalat hírnevét. Ebben a szakaszban kerülnek meghatározásra az együtthatók, hogy elrejtsék a túlbecsült mennyiségeket vagy anyag- és munkadíjakat.

Ennek eredményeképpen az építőiparban működő vállalatok hatékonyságának és jövedelmezőségének növelésére szolgáló fő rendszer nem a döntéshozatali folyamatok automatizálása és felgyorsítása, hanem az anyagok és munkák áráival való spekuláció (5.4-5. ábra). A munkák és anyagok költségeinek túlértékelése a zárt ERP /PMIS - rendszerekben a szürke könyvelés révén történik, az anyagok vagy a munkák mennyiségének átlagos piaci árai feletti százalékok felduzzasztásával, olyan együtthatók segítségével (5.1-6. ábra), amelyeket a "Számítások összeállítása és a munkák költségének kiszámítása az erőforrásbázis alapján" című fejezetben tárgyaltunk.

Ennek eredményeként az ügyfél olyan számítást kap, amely nem a valós költségeket vagy a munka terjedelmét tükrözi, hanem számos rejtett belső együttható származéka. Ugyanakkor az alvállalkozók, hogy megpróbálják teljesíteni a fővállalkozó által meghatározott, alulbecsült árakat, gyakran kénytelenek olcsóbb és rosszabb minőségű anyagokat vásárolni, ami rontja az építés végső minőségét.

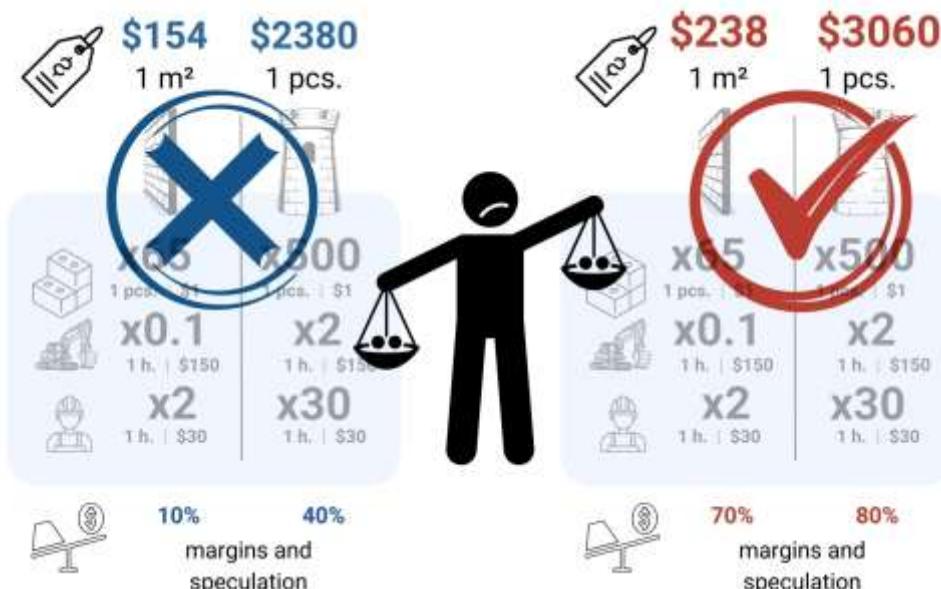
A spekulatív folyamat, amelynek során a profitot a semmiből keresik, végül minden az ügyfeleknek árt, akik megbízhatatlan adatokat kapnak, mind pedig a vezetőknek, akik kénytelenek egyre több spekulációs modellt találni

Ennek eredményeképpen minél nagyobb a projekt, annál nagyobb a bürokrácia az adatok és a folyamatok kezelésében. minden lépés és minden modul gyakran átláthatatlan együtthatókat és felárakat rejt a számítási algoritmusokba és belső eljárásokba ágyazva. Ez nemcsak az ellenőrzést nehezíti, hanem jelentősen torzítja a projekt pénzügyi képét is. A nagy építkezési projekteknél az ilyen gyakorlatok gyakran a végső költségek többszörösére (néha akár tízszeresére) növekedéséhez vezetnek, miközben a valós volumenek és költségek a megrendelő tényleges ellenőrzésén kívül maradnak (2.1-3. ábra A németországi

nagy infrastrukturális projektek tervezett és tényleges költségeinek összehasonlítása.

A McKinsey & Company Imagining the Digital Future of Construction (2016) című jelentése szerint a nagy építési projektek átlagosan 20%-kal később fejeződnek be a tervezettnél, és akár 80%-kal a tervezett költségvetés felett [107].

A becslési és költségvetés-tervezési osztályok a vállalaton belüli legjobban őrzött láncszemekké válnak. A hozzájuk való hozzáférés még a belső szakemberek számára is szigorúan korlátozott, és a zárt logikai és adatbázis-struktúrák miatt lehetetlen torzítások nélkül objektíven értékelni a projektdöntések hatékonyságát. Az átláthatóság hiánya ahhoz vezet, hogy a vállalatok nem a folyamatok optimalizálására kényszerülnek, hanem arra, hogy a számok és tényezők "kreatív" kezelésével küzdjenek a túlélésért (5.3-1. ábra, 5.1-6. ábra - például a "Bid. Factor" paraméter).



Ábra: 5.4-5 A települési szintű spekulációs arányok a vállalatok fő haszna és a munka minősége és a hírnév közötti zsonglórkodés művészete.

Mindez kétséges sé teszi a zárt ERP/PMIS -rendszerek további használatát az építőiparban. A digitális átalakulás és a megrendelők részéről az átláthatóság iránti növekvő igények (10.2-3. ábra) összefüggésében nem valószínű, hogy a projektek megvalósítása hosszú távon továbbra is a rugalmasságot korlátozó, az integrációt akadályozó és az üzleti fejlődést gátló, védett megoldásoktól fog függeni.

És bármennyire is kedvező az építőipari vállalatok számára, hogy zárt adatbázisokban lévő adatsílokkel és átláthatatlan adatokkal dolgoznak - elkerülhetetlen, hogy az építőipar jövője a nyílt platformokra, a gépileg olvasható és átlátható adatstruktúrára és a bizalmon alapuló automatizálásra való áttérést jelentse. Ezt az átalakulást felülről fogják irányítani - az ügyfelek, a szabályozó hatóságok és a társadalom nyomására, akik egyre inkább megkövetelik az elszámoltathatóságot, a fenntarthatóságot, az átláthatóságot és a gazdasági megvalósíthatóságot.

A zárt ERP /PMIS korszakának vége: az építőiparnak új megközelítésekre van szüksége.

A több tízmillió sornyi kódból álló, terjedelmes, moduláris ERP/PMIS -rendszer használata rendkívül megnehezíti a bennük végrehajtott változtatásokat. Ebben az esetben az új platformra való áttérés a vállalat számára már előre konfigurált modulok, az erőforrás-adatbázisokban található több tízezer cikk (5.1-3. ábra) és több ezer kész számítás (5.1-6. ábra) jelenlétében költséges és hosszadalmas folyamatot válik. Minél több kód és örökölt architektúra - annál magasabb a belső hatékonysági szint, és minden egyes új projekt csak ront a helyzeten. Sok vállalatnál az adatmigráció és az új megoldások integrálása többével eposzokká válik, amelyet állandó utómunka és végletes kompromisszumkeresés kísér. Az eredmény gyakran a régi, megszokott platformokhoz való visszatérés, azok korlátai ellenére.

Amint azt az építőipari adatkezelés rendszerszintű hibáiról szóló német Black Book jelentés [108] is kiemeli, az információk szétvoredezetsége és a központosított megközelítés hiánya a hatékonyság hiányának egyik fő oka. Szabványosítás és integráció nélkül az adatok elveszítik értéküket, és inkább archívummá válnak, mint irányítási eszközzé.

Az adatminőség csökkenésének egyik fő oka az építési projektek nem megfelelő tervezése és ellenőrzése, ami gyakran jelentős költségnövekedéshez vezet. A Fekete könyv "Fókusz: A költségrobbanás" című fejezete elemzi azokat a legfontosabb tényezőket, amelyek hozzájárulnak e nemkívánatos következményekhez. Ezek közé tartozik a nem megfelelő igényelemzés, a megvalósíthatósági tanulmányok hiánya és a koordinálatlan tervezés, amely elkerülhető többletköltségekhez vezet.

Egy vállalat érett IT-ökoszisztémájában egy elavult rendszer cseréje ahhoz hasonlítható, mintha egy már felépített épületben egy tartóoszlopot cserélnénk ki. Nem elég egyszerűen eltávolítani a régit, és újat telepíteni - fontos, hogy ezt úgy tegyük, hogy az épület stabil maradjon, a mennyezet ne omoljon be, és minden kommunikáció továbbra is működjön. Itt rejlik a nehézség: bármilyen hiba súlyos következményekkel járhat az egész vállalati rendszerre nézve.

Ennek ellenére az építőipari nagy ERP-termékek fejlesztői továbbra is a megírt kód mennyiséget használják a platformjuk mellett szóló érvként. Szakmai konferenciákon még mindig hallani olyan mondatokat, mint: "150 emberévre telne egy ilyen rendszer újraalkotása", annak ellenére, hogy az ilyen rendszerek funkcionálisának nagy része adatbázisokat és a táblázatokkal való munka meglehetősen egyszerű funkcióit rejt, egy speciális, rögzített felhasználói felületbe csomagolva. A gyakorlatban a "150 emberév" kódmennyiséggel inkább teherré válik, mint versenyelőnyé. Minél több a kód - annál magasabbak a támogatási költségek, annál nehezebb az új feltételekhez való alkalmazkodás, és annál magasabb a belépési küszöb az új fejlesztők és ügyfelek számára.

Sok moduláris építési rendszer ma nehézkes és elavult "Frankenstein-konstrukcióra" hasonlít, ahol bármilyen óvatlan változtatás meghibásodáshoz vezethet. minden egyes új modul tovább bonyolítja az amúgy is túlterhelt rendszert, olyan labirintussá változtatva azt, amelyet csak néhány szakember képes megérteni, és így még nehezebbé teszi a karbantartást és a korszerűsítést.

A komplexitást maguk a fejlesztők is érzékelik, akik időnként szünetet tartanak a refaktorálásban - az architektúra felülvizsgálatában, hogy figyelembe vegyék az új technológiák megjelenését. A komplexitás azonban még akkor is elkerülhetetlenül növekszik, ha a refaktorálásra rendszeresen sor kerül. Az ilyen rendszerek tervezői hozzájárulnak a növekvő komplexitáshoz, de az új felhasználók és szakemberek számára ez leküzdhetetlen akadályá válik. Ennek eredményeképpen minden szakértelem néhány fejlesztő kezében összpontosul, és a rendszer megszűnik skálázhatóvá válni. Rövid távon az ilyen szakértők hasznosak, de hosszú távon a probléma részévé válnak.

A szervezetek továbbra is integrálni fogják a "kis" adatokat a nagy adatokkal, és ostobaság lenne bárki számára azt hinni, hogy egyetlen alkalmazás - legyen az bármilyen drága vagy robusztus - minden képes kezelni [109].

- Phil Simon, a Conversations About Collaboration podcast házigazdája

Jogosan merül fel a kérdés: valóban szükségünk van-e ilyen nehézkes és zárt rendszerekre a munka költségének és időzítésének táblázatos formában történő kiszámításához, ha más iparágak már régóta képesek hasonló feladatokkal megbirkózni, nyílt adatokkal és átlátható logikával rendelkező analitikai eszközökkel?

Jelenleg a zárt moduláris platformok még mindig keresettek az építőiparban, elsősorban a költségszámítás sajátosságai miatt (5.1-7. ábra). Az ilyen rendszereket gyakran szürke vagy átláthatatlan rendszerek működtetésére használják, lehetővé téve a valós költségek elrejtését a megrendelő elől. Ahogy azonban az iparág digitálisan érik, elsősorban a megrendelők, és az úgynevezett "Uberizált korszakba" lép, a közvetítők, nevezetesen az ERP-jükkel rendelkező építőipari vállalatok elveszítik jelentőségüket az idő- és költségszámításokban. Ez örökre meg fogja változtatni az építőipar arcát. Bővebben a könyv utolsó részében és az "Építőipar 5.0: Hogyan keressünk pénzt, ha már nem tudunk elbújni" című fejezetben olvashat.

Az elmúlt 30 év során felhalmozott több ezer örökölt megoldás, amelyek fejlesztésébe több ezer emberévet fektettek, gyorsan eltűnik. A nyílt, átlátható és rugalmas adatkezelésre való áttérés elkerülhetetlen. A kérdés csak az, hogy mely vállalatok lesznek képesek alkalmazkodni ezekhez a változásokhoz, és melyek maradnak a régi modell túszai.

Hasonló helyzet figyelhető meg a CAD (BIM-) eszközök területén, amelyek adatai ma az ERP/PMIS-rendszerekben a tervezési egységek térfogati paramétereit töltik ki. Kezdetben a BIM (még 2002-ben [110] kidolgozott) ötlete egyetlen integrált adatbázis koncepcióján alapult, de a gyakorlatban a BIM-mel való munka ma már egy sor speciális szoftver és formátum egészét igényli. Ami a tervezés és az építésirányítás egyszerűsítésének volt hivatott, az mára a saját fejlesztésű megoldások egy újabb rétegévé vált, amely bonyolítja az integrációt és csökkenti az üzleti rugalmasságot.

Következő lépések: a projektadatok hatékony felhasználása

Ebben a részben bemutattuk, hogyan válnak a strukturált adatok az építési projektek pontos költség- és ütemterv-számításainak alapjává. A QTO, az ütemezési és becslési folyamatok automatizálása csökkenti

a munkaerőköltségeket és jelentősen javítja az eredmények pontosságát.

E rész összegzéseként érdemes kiemelni a főbb gyakorlati lépéseket, amelyek segítségével a tárgyalt megközelítéseket alkalmazhatja a napi feladatai során. Ezek a megközelítések univerzálisak - egyaránt hasznosak egy vállalat digitális átalakításában és a számításokban részt vevő szakemberek minden nap munkájában:

■ Rutinszámítások automatizálása

- Próbáljon meg olyan szabványos munkaköltségeket találni, amelyekhez a munkája során viszonyítani tud.
- Elemezze, hogy az Ön országában milyen módszereket használnak az építkezéseken a munkálatok vagy folyamatok költségének vagy számításának megállapítására (5.1-7. ábra).
- Ha CAD rendszerrel dolgozik - ismerje meg a specifikációk és a QTO adatok automatikus kinyerésének funkciót a CAD (BIM-) szoftverben.
- Az LLM segítségével írjon kódtervezetet a számítások automatizálására.

■ Saját eszközök fejlesztése a QTO számára

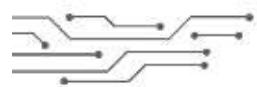
- Szkriptek vagy táblázatok létrehozása a kötetszámlálás automatizálására
- A kategóriák és elemcsoportok szabványosítása az értékelés következetes megközelítése érdekében
- A számítási módszertan dokumentálása az eredmények reprodukálhatóságának biztosítása érdekében az új projektekben.

■ A projekt különböző aspektusainak integrálása a munkádba

- Ha moduláris rendszerekkel dolgozik, próbálja meg a folyamatokat nem csak diagramok vagy grafikonok formájában, hanem az adatok szintjén is megjeleníteni - különösen táblázatok formájában.
- A CAD-adatbázisokból kinyert adatok automatikus összevonásának elsajátítása számításokkal - Python kóddal, csoportosítás, szűrés és aggregálás segítségével.
- A QTO-csoportok egyértelmű vizualizációjának létrehozása a kollégák és az ügyfelek számára történő komplex információk bemutatására.

Ezek a lépések segítenek egy fenntartható, automatizáláson és az adatok szabványosításán alapuló számítási rendszer kiépítésében. Ez a megközelítés javítani fogja a pontosságot és csökkenteni fogja a napi számítási problémák rutinját.

A következő fejezetek a CAD - (BIM-) termékek technikai aspektusaira és azokra az okokra összpontosítanak, amelyek miatt a CAD-adatbázisokat még mindig nehéz integrálni a vállalatok üzleti folyamataiba. Ha Önt most nem érdekli a BIM építőipari bevezetésének története, a CAD-eszközök fejlődése és az ezekkel a technológiákkal való munka technikai vonatkozásai, akkor egyenesen a könyv hetedik, "Adatvezérelt döntéshozatal" című részéhez térhet át.



MAXIMÁLIS KÉNYELEM A NYOMTATOTT VERZIÓVAL

Ön birtokában van a **Data-Driven Construction** ingyenes digitális verziója. A kényelmesebb munkavégzés és az anyagokhoz való gyors hozzáférés érdekében javasoljuk, hogy figyeljen a nyomtatott kiadásra:



Mindig kéznél van: a nyomtatott formátumú könyv megbízható munkaeszközévé válik, amely lehetővé teszi, hogy bármilyen munkaszituációban gyorsan megtalálja és használja a megfelelő ábrákat és diagramokat.

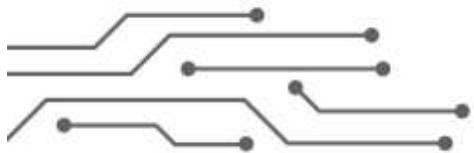
Kiváló minőségű illusztrációk: a nyomtatott kiadásban szereplő összes kép és grafika maximális minőségen jelenik meg.

Gyors hozzáférés az információkhoz: könnyű navigáció, jegyzetek, könyvjelzők készítése és a könyvvel való munka bárhol.

A könyv teljes nyomtatott változatának megvásárlásával egy kényelmes eszközt kap az információkkal való kényelmes és hatékony munkavégzéshez: a képi anyagok gyors használatának lehetőségét a minden nap feladatai között, a szükséges sémkák gyors megtalálását és jegyzetelését. Emellett vásárlásával támogatja a nyílt tudás terjesztését.

Rendelje meg a könyv nyomtatott változatát a következő címen:

datadrivenconstruction.io/books



VI RÉSZ

CAD ÉS BIM: MARKETING, VALÓSÁG ÉS A TERVEZÉSI ADATOK JÖVŐJE AZ ÉPÍTŐIPARBAN

A könyv hatodik része a CAD és a BIM -technológiák fejlődésének és az építőipari adatkezelési folyamatokra gyakorolt hatásuknak kritikai elemzését mutatja be. Nyomon követi a BIM koncepció történelmi átalakulását az integrált adatbázis eredeti ötletétől a szoftvergyártók által támogatott jelenlegi marketingkonstrukcióig. Értékeli a saját formátumoknak és a zárt rendszereknek a projektadatok kezelésének hatékonyságára és az építőipar általános teljesítményére gyakorolt hatását. Részletesen elemzik a különböző CAD-rendszerek kompatibilitásának problémáit és az építőipari vállalatok üzleti folyamataival való integrációjuk nehézségeit. Az egyszerűsített nyílt adatformátumok, például az USD felé mutató jelenlegi tendenciák és az ágazatra gyakorolt lehetséges hatásuk megvitatásra kerülnek. A zárt rendszerekből történő információszerzés alternatív megközelítéseit mutatják be, beleértve a reverse engineering technikákat. A mesterséges intelligencia és a gépi tanulás alkalmazásának kilátásait elemzik a tervezés és az adatelemzési folyamatok automatizálása érdekében az építőiparban. A szoftvergyártók érdekei helyett a felhasználók valós igényeihez igazodó tervezési technológiák fejlesztésére vonatkozó előrejelzéseket fogalmaznak meg.

FEJEZET 6.1.

A BIM- KONCEPCIÓK MEGJELENÉSE AZ ÉPÍTŐIPARBAN

Eredetileg ez a hatodik rész, amely a CAD (BIM) témakörének szentelt, nem szerepelt a könyv első változatában. A szabadalmaztatott formátumok, a geometriai magok és a zárt rendszerek témái túlságosan technikai jellegűek, túlterheltek a részletekkel, és látszólag haszontalanak tűnnek azok számára, akik csak azt szeretnék megérteni, hogyan kell az adatokkal dolgozni. A visszajelzések és a könyv első változatának pontossására irányuló kérések azonban azt mutatták, hogy a CAD-rendszerek belső működésének, a geometriai kernelek, a különböző formátumok és az ugyanazon adatokhoz tartozó inkompatibilis tárolási rendszerek sokféleségének megértése nélkül lehetetlen igazán megérteni, hogy a gyártók által támogatott koncepciók miért nehezítik gyakran az információkkal való munkát, és miért akadályozzák a nyílt parametrikus tervezésre való áttérést. Ezért ez a rész a könyv szerkezetében külön helyet kapott. Ha a CAD (BIM) nem prioritás az Ön számára, akkor egyenesen a következő részre ugorhat: "VII. RÉSZ: Adatvezérelt döntéshozatal, analitika, automatizálás és gépi tanulás".

A BIM és a nyílt BIM mint a CAD- gyártók marketingkoncepcióinak kialakulásának története.

A digitális adatok 1990-es évekbeli megjelenésével a számítástechnika nemcsak az üzleti folyamatokban, hanem a tervezési folyamatokban is megjelent, és olyan fogalmakhoz vezetett, mint a CAD (számítógépes tervezési rendszerek), majd később a BIM (épületinformációs modellezés)

Azonban, mint minden innováció, ezek sem jelentik a fejlődés végpontját. Az olyan koncepciók, mint a BIM fontos mérföldkőnek számítanak az építőipar történetében, de előbb-utóbb jobb eszközöknek és megközelítéseknek adhatják át a helyüket, amelyek jobban megfelelnek a jövő kihívásainak.

A CAD gyártók befolyásától túlterhelt és a saját megvalósításának összetettsége miatt összeavarodott BIM koncepciója, amely 2002-ben jelent meg, talán nem éri meg harmincadik születésnapját, mint egy rocksztár, amely fényesen felcsillant, de gyorsan elhalványult. Az ok egyszerű: az adattudósok igényei gyorsabban változnak, mint ahogy a CAD-gyártók alkalmazkodni tudnak hozzájuk.

A minőségi adatok hiányával szembesülve a mai építőipari szakemberek a CAD- projektekből származó nyílt adatokhoz való hozzáférést és a platformok közötti átjárhatóságot követelik meg, hogy egyszerűsítsék elemzésüket és feldolgozásukat. A CAD-adatok összetettsége és a CAD-adatok zavaros feldolgozása negatív hatással van az építési folyamat minden résztvevőjére: a tervezőkre, a projektmenedzserekre, a helyszínen dolgozó építőmunkásokra és végső soron az ügyfélre.

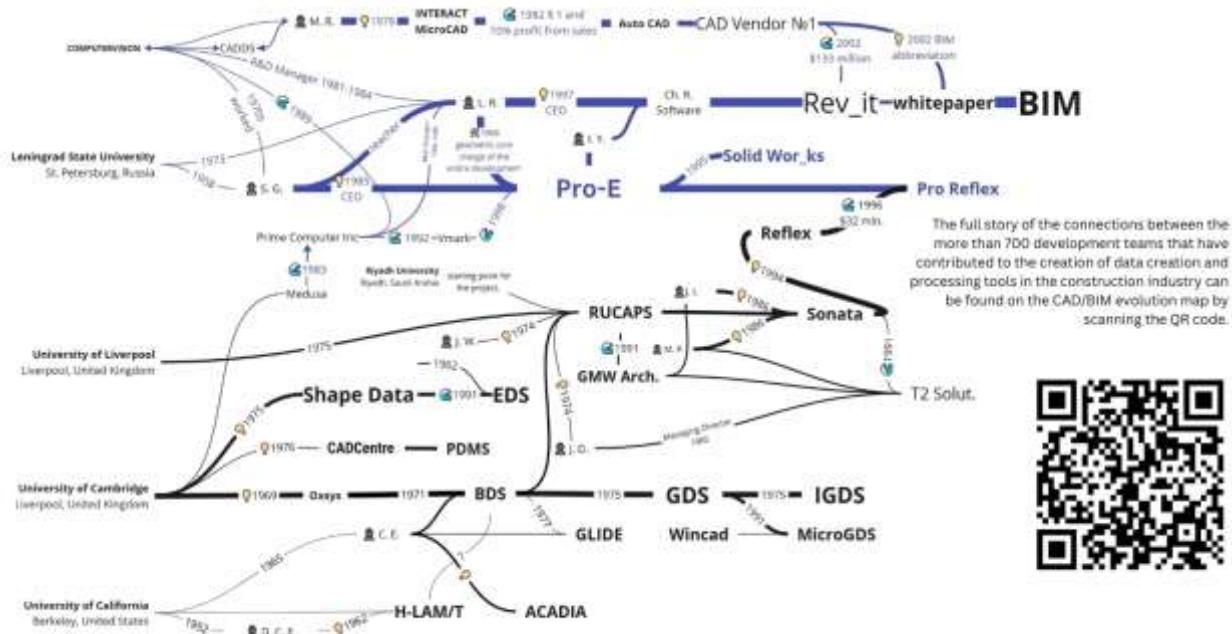
Ma a teljes értékű üzemeltetési adatkészlet helyett a megrendelő és a beruházó CAD- formátumban kapja a konténereket, amelyekhez bonyolult geometriai magok, az adatsémák megértése, évente frissített API - dokumentáció és speciális CAD szoftver (BIM) szükséges az adatokkal való munkához. Ugyanakkor a tervezési adatok nagy része kihasználatlanul marad.

A mai tervezési és építési világban a CAD-adatokhoz való hozzáférés bonyolultsága a projektmenedzsment

túlméretezéséhez vezet. A CAD-adatokkal dolgozó vagy BIM-megoldásokat fejlesztő közép- és nagyvállalatok vagy kénytelenek szoros kapcsolatot fenntartani a CAD-szállítókkal megoldásokkal, hogy az adatokhoz API-kon keresztül férjenek hozzá, vagy a CAD-szállítók korlátozásainak megkerülésével drága SDK-konvertálók használatával, hogy visszafejtésük a, hogy nyílt adatokhoz jussanak [75].

A saját adatokra vonatkozó megközelítés elavult, és már nem felel meg a mai digitális környezet követelményeinek. A jövő két típusra fogja osztani a vállalatokat: azokra, amelyek hatékonyan használják a nyílt adatokat, és azokra, amelyek elhagyják a piacot.

A BIM (Building Information Modeling) fogalma az építőiparban az egyik nagy CAD-gyártó 2002-es Whitepaper BIM [54] című kiadványával jelent meg az építőiparban, és a BOM (Bills of Materials) gépészeti fogalmával kiegészülve a projektadatok létrehozásának és feldolgozásának parametrikus megközelítéséből származik (6.1-1. ábra). A tervadatok létrehozásának és feldolgozásának parametrikus megközelítése az elsők között valósult meg a gépészeti tervezésre szolgáló Pro-E rendszerben (MCAD). Ez a rendszer lett számos modern CAD[111], beleértve a ma az építőiparban használtakat is.



6.1-1 ábra A BIM-koncepció és a hasonló koncepciók történetének térképe.

Az újságírók és AEC-tanácsadók, akik a 2000-es évek elejéig a CAD-eszközöket népszerűsítették - szolgáltatók, 2002-től a Whitepaper BIM-re fordították figyelmüket. A Whitepaper BIM 2002-2004 és a 2002-ben, 2003-ban, 2005-ben és 2007-ben megjelent cikkek kulcsszerepet játszottak a BIM koncepció népszerűsítésében az építőiparban [112].

Az épületinformációs modellezés egy stratégia..... [CAD-szállító cég neve] az információs technológia alkalmazására az építőiparban.

- BIM Whitepaper, 2002 [60].

A 2000-es évek közepére a "kutatók" elkezdték összekapcsolni a CAD- gyártó által 2002-ben közzétett BIM-koncepciót korábbi tudományos munkákkal, például Charles Eastman BDS című munkájával, amely olyan rendszerek alapjává vált, mint a GLIDE, GBM, BPM, RUCAPS. Charles Eastman úttörő Building Description System (1974) című munkájában lefektette a modern információs modellezés elméleti alapjait. Az "adatbázis" kifejezés 43 alkalommal fordul elő munkájában (6.1-2. ábra) - az "épület" szó kivételével minden másnál gyakrabban.

Eastman fő gondolata az volt, hogy egy épületre vonatkozó összes információt - a geometriától az elemek tulajdonságaiig és azok kapcsolataiig - egyetlen strukturált adatbázisban kell tárolni. Ebből az adatbázisból automatikusan generálhatók és elemezhetők a rajzok, a specifikációk, a számítások és a szabályoknak való megfelelés. Eastman kifejezetten bírálta a rajzokat, mint elavult és felesleges kommunikációs módszert, rámutatva az információk megkettőződésére, a frissítéssel kapcsolatos problémákra és a kézi frissítések szükségességére, amikor változások történnek. Ehelyett egyetlen digitális modellt javasolt egy adatbázisban, ahol minden változtatás egyszer történik, és automatikusan megjelenik az összes nézetben.

Figyelemre méltó, hogy Eastman koncepciójában nem a vizualizációt állította a középpontba. Rendszerének középpontjában az információ állt: paraméterek, kapcsolatok, attribútumok, elemzési és automatizálási lehetőségek. A rajzok az ó felfogásában csak az adatbázisból származó adatok megjelenítésének egyik formája voltak, nem pedig a tervezési információk elsődleges forrása.

A vezető CAD szállítójának első BIM Whitepaperében a "database" kifejezés ugyanolyan gyakran szerepelt, mint Charles Eastman BDS-ében - hét oldalon keresztül 23-szor [60], és az "Building", "Information", "Modeling" és "Design" után az egyik legnépszerűbb szó volt a dokumentumban. 2003-ra azonban az "adatbázis" kifejezés csak kétszer fordul elő hasonló dokumentumokban [61], és a 2000-es évek végére az adatbázisok téma gyakorlatilag eltűnt a tervezési adatokról szóló vitából. Ennek eredményeképpen az "egyetlen integrált adatbázis a vizuális és mennyiségi elemzéshez" koncepciója soha nem valósult meg teljes mértékben.

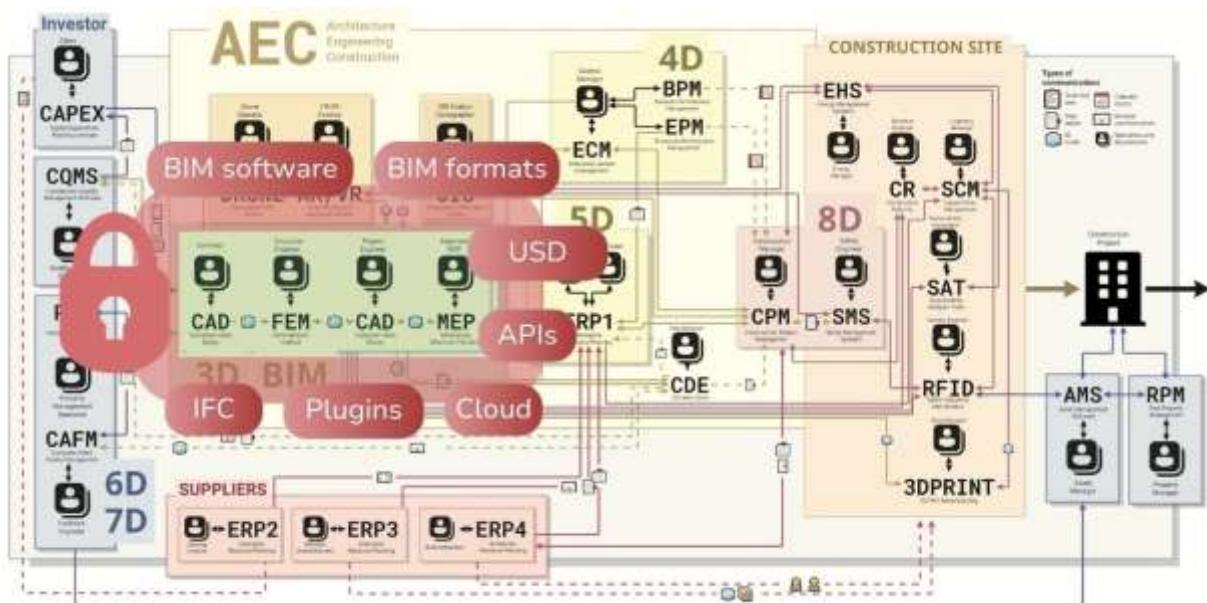
Így az építőipar Charles Eastman progresszív BDS-koncepciójától, amely az adatbázisokra helyezte a hangsúlyt, és Samuel Geisberg elképzéseitől a tervezési adatok adatbázisokból történő automatikus frissítéséről a Pro-E gépészeti terméken (a ma az építőiparban használt népszerű CAD -megoldások elődje) a jelenleg forgalmazott BIM -ig jutott el, ahol az adatbázisokon keresztüli adatkezelésről alig esik szó, annak ellenére, hogy ez volt az eredeti elméleti koncepció mögött álló elképzélés.

Az adatok az eladóké, nem a felhasználóké. A projektinformációkat szabadalmaztatott formátumokba vagy felhőszolgáltatásokba zárják, ahelyett, hogy nyílt és független formátumokban lennének elérhetők.

A tervezőmérnökök és a projektmenedzserek gyakran nem férnek hozzá a CAD-adatbázisokhoz, sem ahhoz a formátumhoz, amelyben a saját projektadataikat tárolják. Ez lehetetlenné teszi az információk gyors ellenőrzését vagy az adatszerkezetre és -minőségre vonatkozó követelmények megfogalmazását (6.1-3. ábra). Az ilyen adatokhoz való hozzáféréshez API-kon és beépülő modulokon keresztül összekapcsolt speciális programok egész sorára van szükség, ami az építőiparban a folyamatok túlzott bürokratizálásához vezet. Eközben ezeket az adatokat egyszerre több tucat információs rendszer és több száz szakember használja.

Képesnek kell lennünk arra, hogy mindezeket az adatokat [CAD (BIM)] digitálisan tároljuk, és életciklus- és folyamatmenedzsment-szoftvereket értékesítsünk, mert minden mérnökre [tervezőre], aki létrehoz valamit [CAD-szoftverben], tíz ember jut, aki ezekkel az adatokkal dolgozik" [41].

- A CAD - a BIM koncepciót megalkotó, 2005 - vezérigazgatója.



Ábra 6.1-3 CAD- (BIM-) adatbázisok továbbra is az egyik utolsó zárt rendszer az informatikai osztályok és az adatkezelők számára az építőipari üzleti ökoszisztémában.

Amikor nyilvánvalóvá válik, hogy a BIM inkább az adatbázisok kereskedelmi forgalomba hozatalának eszköze, mintsem egy teljes értékű adatbázis-kezelő eszköz, felmerül a logikus kérdés: hogyan nyerhetjük vissza az adatok feletti ellenőrzést? A válasz a nyílt adatstruktúrák használata, ahol a felhasználó, nem pedig a szoftvergyártó lesz az információ tulajdonosa.

Az építőipari felhasználók és megoldásfejlesztők, akárcsak más iparágakban dolgozó társaik, elkerülhetetlenül eltávolodnak az elmúlt 30 évben uralkodó, homályos szoftvergyártói terminológiától, és a digitalizáció kulcsfontosságú aspektusaira - az "adatokra" és a "folyamatokra" - összpontosítanak.

Az 1980-as évek végén az építőiparban a digitális fejlesztés kulcsfontosságú területét az adathozzáférés és a projektinformációk kezelése kérdéseként mutatták be. Idővel azonban a hangsúly eltolódott. Az adatokkal való munkavégzés átlátható és hozzáférhető megközelítéseinek kidolgozása helyett az IFC formátumot és a nyílt BIM koncepciót aktívan támogatták, mint olyan kísérleteket, amelyek a szakemberek figyelmét a projektadatbázis-kezelés témáiról próbálták elterelni.

A nyílt formátumú IFC megjelenése az építőiparban

Az úgynevezett nyílt formátum IFC (Industry Foundation Classes) szabvánnyként van elhelyezve, hogy biztosítsa a különböző CAD (BIM-) rendszerek közötti átváthatóságot. Fejlesztése a nagy CAD-gyártók által létrehozott és ellenőrzött szervezetek keretein belül történt. Az IFC formátum alapján két CAD- cég 2012-ben kidolgozta az OPEN BIM[63] marketingkoncepcióját.

Az IFC (Industry Foundation Classes) egy nyílt szabvány az építőiparban történő adatcserére, amelyet a különböző CAD - (BIM-) rendszerek közötti átváthatóság biztosítására terveztek.

A nyílt BIM - koncepció a CAD-adatbázisokból származó információkkal való munkát és a rendszerek közötti információcserét a CAD-adatok cseréjére szolgáló nyílt formátumon - IFC - keresztül foglalja magában.

Az Open BIM Program egy marketing kampány, amelyet... [1 CAD gyártó],... [2 CAD gyártó] és más vállalatok kezdeményeztek, hogy ösztönözzék és megkönnyítsék az OPEN BIM koncepció globális, összehangolt népszerűsítését az AEC iparágban, a program résztvevői számára elérhető egységes kommunikációval és közös márkajelzéssel.

- A CAD-szállító weboldaláról, OPEN BIM program, 2012 [113].

Az IFC-t a Müncheni Műszaki Egyetem adaptálta a STEP gépészeti formátumból az 1980-as évek végén, majd később egy nagy tervező cég és egy nagy CAD- gyártó bejegyeztette, és 1994-ben megalakította az IAI-t (Industry Alliance for Interoperability) [114] (6.1-4. ábra). Az IFC formátumot a különböző CAD rendszerek közötti átváthatóságra tervezte, és a STEP gépészeti formátumban lefektetett elveken alapult, amely viszont a CAD felhasználók és gyártók egy csoportja által 1979-ben a NIST (The National Institute of Standards and Technology) és az USA védelmi minisztériumának támogatásával létrehozott IGES formátumból alakult ki [115].

Az IFC összetett szerkezete, a geometriai magtól való szoros függése, valamint a formátum különböző szoftvermegoldások általi megvalósításának eltérései azonban számos problémát okoztak a gyakorlati alkalmazásban. Hasonló nehézségekkel - a részletesség elvesztése, a pontosság korlátozása és a köztes formátumok használatának szükségessége - korábban a gépészeti szakemberek is találkoztak, amikor az IGES és STEP formátumokkal dolgoztak, amelyekből az IFC kialakult.



Ábra 6.1-4 A fejlesztőcsapatok és a CAD-termékek közötti kapcsolatok térképe (BIM) [116].

2000-ben ugyanaz a CAD -szolgáltató, amely az IFC formátumot regisztrálta és létrehozta az IAI (később bS) szervezetet, kiadja az "Integrated Design and Manufacturing: Benefits and Rationale" című fehér könyvet. [65]. A dokumentum hangsúlyozza annak fontosságát, hogy az egyazon rendszeren belüli programok közötti adatcsere során a teljes adatgranuláció megtartása fontos, anélkül, hogy semleges formátumokat használnának, mint például az IGES, STEP [azonos az IFC-vel]. Ehelyett azt javasolták, hogy az alkalmazásoknak közvetlen hozzáféréssel kell rendelkezniük a mögöttes CAD-adatbázishoz, hogy elkerüljék az információk pontosságának elvesztését.

2002-ben ugyanez a CAD gyártó vásárolja meg a parametrikus BOM terméket (3.1-18. ábra, további részletek a harmadik részben), és ennek alapján alakítja ki a BIM koncepciót. Ennek eredményeképpen az építési projektadatok cseréjére csak CAD-formátumokat vagy az IFC formátumot (STEP) használják, amelynek korlátairól 2000-ben maga a CAD-gyártó írt, aki ezt a formátumot az építőiparban bevezette.

Az építési adatok létrehozására és feldolgozására szolgáló eszközök létrehozásában részt vevő több mint 700 fejlesztőcsapat interakciójának részletes történetét a "A CAD (BIM) fejlődése" című térkép mutatja be [116]. [116].

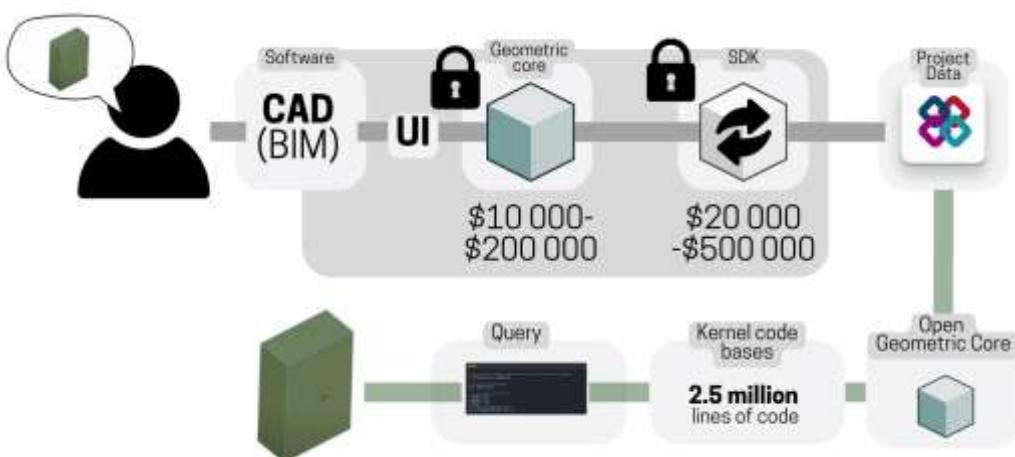
A nyílt formájú IFC a tervezési elemek geometriai leírásából és a metainformációk leírásából áll. A geometria IFC formátumban történő ábrázolására különböző módszereket használnak, például a CSG és a Swept Solids: a parametrikus ábrázolás BREP azonban az elemek geometriájának IFC formátumban történő

átvitelének vezető szabványává vált, mivel ez a formátum támogatott a CAD- (BIM-) programokból történő exportáláskor, és lehetővé teszi az elemek lehetséges szerkesztését az IFC CAD programokba történő visszaimportálásakor.

IFC formátum probléma a geometriai magtól függően

A legtöbb esetben, amikor az IFC weboldalon a geometria parametrikusan van definiálva (BREP), lehetetlenné válik a geometriai tulajdonságok, például a projektelemek térfogatának vagy területének megjelenítése vagy lekérdezése csak egy IFC fájl segítségével, mivel a geometriával való munkához és annak megjelenítéséhez ebben az esetben szükség van egy geometriamagra (6.1-5. ábra), amely kezdetben hiányzik.

A Geometry kernel egy olyan szoftverkomponens, amely alapvető algoritmusokat biztosít a CAD (CAD), BIM és más mérnöki alkalmazásokban a geometriai objektumok létrehozásához, szerkesztéséhez és elemzéséhez. Felelős a 2D és 3D -geometria létrehozásáért és a rajta végzett műveletekért, mint például: Boole-műveletek, simítás, metszéspontok, transzformációk és megjelenítés.

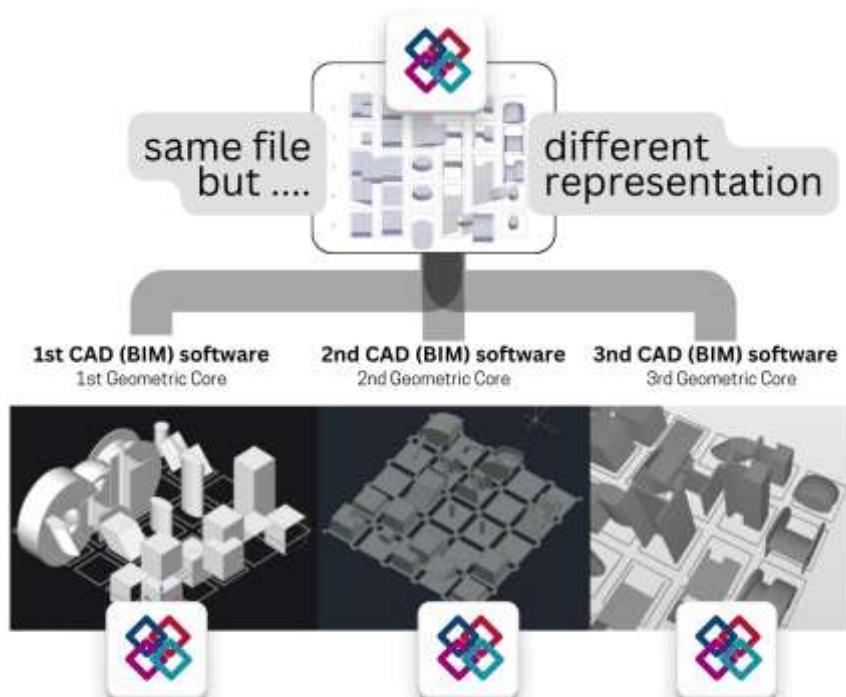


Ábra 6.1-5 A geometria CAD-szoftvereken keresztül történő létrehozása ma már a saját geometriamagokon és SDK-ken keresztül történik, amelyek gyakran nem a CAD-gyártók tulajdonában vannak.

Minden CAD program és minden parametrikus vagy IFC formátummal dolgozó program rendelkezik saját vagy vásárolt geometriai maggal. És ha az IFC-BREP formátumú primitív elemekkel nem lehet probléma, és a különböző geometriai kernelekkel rendelkező programokban ezek az elemek hasonlóan megjeleníthetők, de a geometriai kernelek különböző motorjaival kapcsolatos problémákon kívül vannak olyan elemek is, amelyeknek sajátos sajátosságaik vannak a helyes megjelenítéshez. Ezt a problémát részletesen tárgyalja a 2019-ben megjelent "A reference study of IFC software support" című nemzetközi tanulmány [117].

Ugyanazok a szabványosított adatkészletek ellentmondásos eredményeket hoznak kevés közös mintát találunk, és komoly problémákat találtak a szabvány [IFC] támogatásában, valószínűleg a szabványos adatmodell igen nagyfokú összetettsége miatt. Ebben részben maguk a szabványok is hibásak, mivel gyakran hagynak néhány részletet meghatározatlanul, nagy szabadsági fokokkal és különböző lehetséges értelmezésekkel. Az objektumok szervezésében és tárolásában nagyfokú komplexitást tesznek lehetővé, ami nem kedvez a hatékony egyetemes megértésnek, az egyedi megvalósításoknak és a konzisztens adatmodellezésnek [117].

- Referenciatanulmány az IFC szoftvertámogatásról, 2021



Source: Reference study of IFC software support: the GeoBIM benchmark 2019 — Part I <https://arxiv.org/pdf/2007.10951.pdf>

6.1-6 ábra A különböző geometriai kernelek különböző ábrázolásokat adnak ugyanannak a parametrikusan leírt geometriának ([117] alapján).

Az "egyes rendelkezések" helyes értelmezése az IFC-t kidolgozó speciális szervezetek fizetett tagjai számára elérhető. Következésképpen, aki fontos ismeretekhez akar hozzáférni az IFC bizonyos jellemzőiről, az megpróbál együttműködni a nagy CAD- gyártókkal, vagy saját kutatással jut el a jellemzők minőségi megfontolásához.

Ön belebotlik egy kérdésbe az adatok IFC formátumon keresztül történő importálásával és exportálásával kapcsolatban és megkérdezi a többi forgalmazót: "Miért van így az IFC fájlban információ a helyiségek parametrikus átviteléről? A nyílt specifikáció nem mond erről semmit". Válasz a "hozzáértőbb" európai szállítók részéről: "Igen, nincs benne, de megengedett".

- A CAD 2021 fejlesztőjének interjújából [118].

Az IFC a geometriát parametrikus primitíveken keresztül írja le, de nem tartalmaz beágyazott magot - ennek szerepét a CAD-program tölti be, amely a geometriát a geometriamagon keresztül állítja össze. A geometriamag végzi a matematikai számításokat és határozza meg a metszeteket, míg az IFC csak az adatokat szolgáltatja az értelmezéshez. Ha az IFC hibás felületeket tartalmaz, a különböző geometriamaggal rendelkező programok a kerneltől függően vagy figyelmen kívül hagyják azokat, vagy hibát produkálnak.

Ennek eredményeképpen az IFC formátummal való munkavégzéshez meg kell válaszolni a fő kérdést, amelyre nehéz egyértelmű választ találni - milyen eszközzel, milyen geometriai maggal kell használni ahhoz, hogy a projektben eredetileg a CAD programban, amelyből az IFC-t kaptuk, azt az adatminőséget kapjuk, amellyel a projekt eredetileg rendelkezett?

Az adatminőségi problémák és az IFC formátum bonyolultsága nem teszi lehetővé a tervezési adatok közvetlen felhasználását a folyamatok automatizálására, elemzésére és adatfeldolgozására, ami gyakran arra készíteti a fejlesztőket, hogy elkerülhetetlenül zárt CAD -megoldásokat használjanak, amelyek "minőségi" hozzáférést biztosítanak az adatokhoz[63], amiről maga a gyártó írt, aki 1994-ben regisztrálta az IFC-t [65].

Az IFC paraméterek leképezésének és generálásának minden sajátossága a geometria magban csak olyan nagy fejlesztőcsapatok által valósítható meg, amelyeknek van tapasztalatuk a geometria magokkal való munkában. Ezért az IFC formátum sajátosságainak és bonyolultságának jelenlegi gyakorlata elsősorban a CAD-gyártók számára előnyös, és sok hasonlóságot mutat a nagy szoftvergyártók "adopt, extend, destroy" stratégiájával, amikor a szabvány növekvő bonyolultsága valójában akadályokat állít a kis piaci szereplők elé [94].

nagy gyártók ilyen stratégiája az lehet, hogy a nyílt szabványokat adaptálják, saját fejlesztésű bővítményeket és funkciókat adnak hozzá, hogy a felhasználók függőséget alakítsanak ki a termékeiktől, és így kiszorítsák a versenytársakat.

Az IFC formátum, amelyet a különböző CAD- (BIM-) rendszerek közötti univerzális hídnak szántak, valójában a különböző CAD platformok geometriai magjai közötti kompatibilitási problémák jelzőjeként működik, hasonlóan a STEP formátumhoz, amelyből eredetileg kialakult.

Ennek eredményeképpen ma a IFC ontológia teljes és magas színvonalú megvalósítása a nagy CAD-gyártók számára megvalósítható, akik jelentős erőforrásokat tudnak befektetni az összes entitás támogatásába és

azok saját belső geometriai magjukhoz való hozzárendelésébe, ami az IFC mint szabvány esetében nem létezik. A nagy gyártók emellett képesek egymás között egyeztetni a funkciók technikai részleteit, amelyek még az IFC formátumot fejlesztő szervezetek legaktívabb résztvevői számára sem állnak rendelkezésre.

A kis független csapatok és a nyílt forráskódú projektek számára, amelyek az interoperábilis formátumok fejlesztését igyekeznek támogatni, komoly problémát jelent a házon belüli geometriai mag hiánya. Enélkül gyakorlatilag lehetetlen figyelembe venni a platformok közötti adatcserével kapcsolatos összes különböző finomságot és árnyalatot.

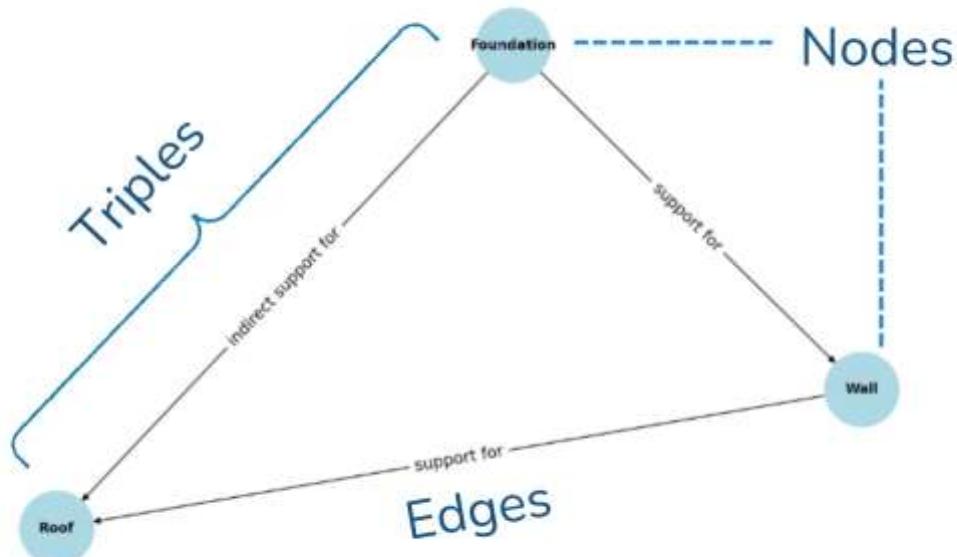
Az IFC parametrikus formátum és a nyílt BIM koncepció kialakulásával az építőiparban felerősödtek a viták az ontológia és a szemantika szerepéiről az adat- és folyamatkezelésben.

Megjelenés a szemantika és az ontológia téma körének felépítésében

A szemantikus internet 1990-es elképzeléseinek és az IFC formátum kifejlesztésében részt vevő szervezetek erőfeszítéseinek köszönhetően, a szemantika és az ontológiák a 2020-as évek közepére az építőiparban tárgyalt szabványosítás egyik kulcsfontosságú elemévé váltak.

A szemantikus technológiák a heterogén adatok nagy tömbjeinek egységesítését, szabványosítását és módosítását, valamint az összetett keresések megvalósítását jelentik.

A szemantikus adatok tárolására az OWL (Web Ontology Language) szolgál, amelyet RDF-gráfkként -triplák (Resource Description Framework) (6.1-7. ábra) reprezentálnak. Az OWL gráf adatmodellekre utal, amelyek típusait az "Adatmodellek: adatkapcsolatok és elemek közötti kapcsolatok" című fejezetben részletesen tárgyaltuk.



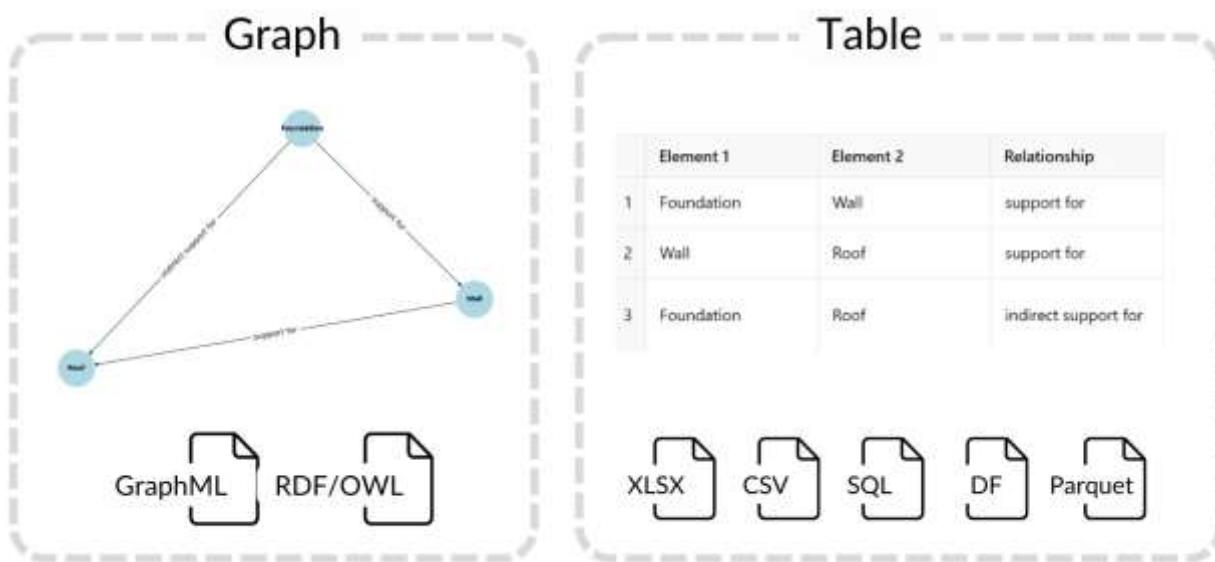
6.1-7. ábra RDF adatmodell: csomópontok, élek és hármasok, amelyek az építőelemek közötti kapcsolatokat szemléltetik.

Elméletileg a risonerek (automatikus logikai következtető szoftverek) logikai következtetése lehetővé teszi új állítások levezetését az ontológiákból. Például, ha a építési ontológia rögzíti, hogy "az alapozás a fal támasza" és "a fal a tető támasza" (6.1-7. ábra), akkor a risoner képes automatikusan levezetni, hogy "az alapozás a tető támasza".

Egy ilyen mechanizmus hasznos az adatelemzés optimalizálásához, mert így elkerülhető az összes függőség explicit előírása. Ugyanakkor nem hoz létre új tudást, hanem csak azonosítja és strukturálja a már ismert tényeket.

A szemantika önmagában nem hoz létre új jelentést vagy tudást, és ebből a szempontból nem jobb, mint más adattárolási és -feldolgozási technológiák. A relációs adatbázisokból származó adatok hármasokként való ábrázolása nem teszi azokat értelmesebbé. A táblázatok gráfszerkezetekkel való helyettesítése hasznos lehet az adatmodellek egységesítése, a könnyű visszakeresés és a biztonságos szerkesztés szempontjából, de nem teszi "okosabbá" az adatokat - a számítógép nem kezdi jobban megérteni a tartalmukat.

Az adatokban lévő logikai kapcsolatok bonyolult szemantikai technológiák nélkül is megszervezhetők (6.1-8. ábra). A hagyományos relációs adatbázisok (SQL), valamint a CSV vagy XLSX formátumok hasonló függőségek kialakítását teszik lehetővé. Például egy oszlopos adatbázisban hozzáadhatunk egy "tető alátámasztása" mezőt, és a fal létrehozásakor automatikusan társíthatjuk a tetőt az alapozással. Ez a megközelítés RDF, OWL, gráfok vagy risonerek használata nélkül valósul meg, egyszerű és hatékony megoldás marad az adatok tárolására és elemzésére.



6.1-8. ábra Az azonos logikai kapcsolatok ábrázolására szolgáló grafikus és táblázatos adatmodellek összehasonlítása.

Számos nagy építőipari vállalat és az IFC formátumot fejlesztő szervezet [94] döntése, hogy követi a szemantikus web koncepcióját, amely az 1990-es évek végén ígéretesnek tűnt, jelentős hatással volt az építőipari szabványok fejlődésére.

A paradoxon azonban az, hogy maga a szemantikus web koncepciója, amelyet eredetileg az internetre szántak, még a saját környezetében sem terjedt el széles körben. Az RDF és az OWL fejlődése ellenére a teljes értékű szemantikus web eredeti elképzelésében nem jelent meg, és létrehozása már most is valószínűtlen.

Miért nem váltják be a szemantikus technológiák az építőiparban a hozzájuk fűzött reményeket?

Más iparágak is szembesültek a szemantika felhasználására szolgáló technológiák korlátaival. A játékiparban a játékobjektumok és interakcióik ontológiák segítségével történő leírására tett kísérletek a változások nagy dinamikája miatt hatástartannak bizonyultak. Ennek eredményeképpen az egyszerűbb adatformátumok, például az XML és a JSON, valamint az algoritmikus megoldások kerültek előtérbe. Hasonló volt a helyzet az ingatlanszektorban is: a terminológia regionális különbségei és a gyakori piaci változások miatt az ontológiák használata túlságosan bonyolultnak bizonyult, míg az egyszerű adatbázisok és az olyan szabványok, mint a RETS [119] jobban megbirkóztak az adatcseré kihívásaival.

A technikai nehézségek, például a jelölés bonyolultsága, a nagy munkaigényű támogatás és a fejlesztők alacsony motivációja lelassította a szemantikus web bevezetését és a gazdaság más ágazataiban. Az RDF (Resource Description Framework) nem vált tömegszabvánnyá, az ontológiák pedig túl bonyolultnak és gazdaságilag indokolatlannak bizonyultak.

Ennek eredményeképpen nem valósult meg az az ambiciózus elképzelés, hogy létrehozzák a globális

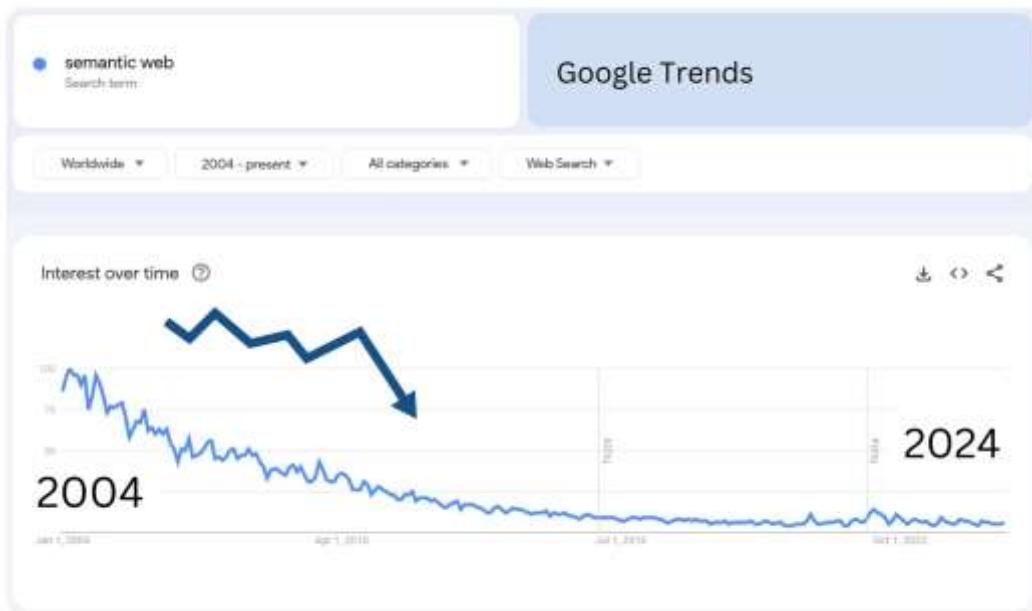
szemantikus webet. Bár a technológia egyes elemei, mint például az ontológiák és a SPARQL, bekerültek a vállalati megoldásokba, az eredeti cél, egy egységes, átfogó adatstruktúra létrehozása nem valósult meg.

Az olyan internet koncepciója, amelyben a számítógépek képesek értelmet adni a tartalomnak, technikailag nehéznek és kereskedelmileg veszteségesnek bizonyult. Ezért az ötletet támogató vállalatok végül az egyes hasznos eszközökre korlátozták a használatát, és az RDF és az OWL inkább a nagyon speciális vállalati igényekre, mint az internet egészére maradt. A Google Trends (6.1-9. ábra) elmúlt 20 évre vonatkozó elemzése azt sugallja, hogy a szemantikus webnek talán nincs több kilátása.

Nincs szükség az entitások szükségtelen megsokszorozására. Ha egy jelenségnak több logikailag konzisztens magyarázata van, amelyek egyformán jól magyarázzák azt, akkor - ha minden más dolog egyenlő - a legegyeszerűbbet kell előnyben részesítenünk.

borotvája

Itt felmerül a logikus kérdés: miért használunk egyáltalán triplákat, emelkedőket és SPARQL építésben, amikor az adatokat népszerű strukturált lekérdezésekkel is feldolgozhatjuk (SQL, Pandas, Apache®)? A vállalati alkalmazásokban az SQL az adatbázisokkal való munka szabványa. A SPARQL ezzel szemben bonyolult gráfstruktúrákat és speciális szoftvereket igényel, és a Google trendjei szerint nem kelti fel a fejlesztők érdeklődését.



6.1-9 A Google statisztikái szerint a "szemantikus internet" lekérdezések iránti érdeklődés.

A gráf adatbázisok és az osztályozási fák bizonyos esetekben hasznosak lehetnek, de alkalmazásuk a legtöbb minden nap feladat esetében nem minden indokolt. Ennek következtében a tudásgráfok létrehozásának és a szemantikus webes technológiák használatának csak akkor van értelme, ha különböző forrásokból származó adatokat kell egységesíteni, vagy összetett logikai következtetéseket kell levonni.

A táblázatokról a grafikus adatmodellekre való áttérés javítja a keresést és egységesíti az információáramlást, de nem teszi az adatokat értelmesebbé a gépek számára. A kérdés nem az, hogy kell-e szemantikus technológiákat használni, hanem az, hogy hol jelentenek valódi különbséget. Mielőtt bevezetné az ontológiát, a szemantikát és a gráfadatbázisokat az Ön vállalatánál, tájékozódjon arról, hogy mely vállalatok alkalmazzák már sikeresen ezeket a technológiákat, és hol vallottak kudarcot.

A nagyra törő várakozások ellenére a szemantikus technológiák soha nem váltak az építőiparban az adatok strukturálására szolgáló általános megoldássá. A gyakorlatban ezek a technológiák nem vezettek univerzális megoldáshoz, hanem csak újabb komplexitásokat eredményeztek, és ezek az erőfeszítések a szemantikus internet koncepciójának meg nem valósult ambícióit tükrözik, ahol a várakozások messze meghaladták a valóságot.

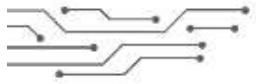


Ábra 6.1-10 Geometria és információ az építési folyamatokban: az összetett CAD és BIM-rendszerektől az egyszerűsített adatokig az analitikához.

Míg az informatikában a szemantikus web kudarcait az új technológiák megjelenése ellensúlyozta (big data, IoT, gépi tanulás, AR/VR), addig az építőiparban nincs ilyen alkalom.

A projektelemek közötti adatkapcsolatok közlésére szolgáló fogalmak használatával kapcsolatos kihívások mellett továbbra is fennáll egy alapvető probléma, mégpedig az adatok elérhetősége. Az építőipart még mindig a zárt rendszerek uralják, ami megnehezíti az adatokkal való munkát, az információk megosztását és a folyamatok hatékonyságának javítását.

Az adatok zárt jellege az egyik legfontosabb akadály, amely akadályozza a digitális megoldások fejlesztését az építőiparban. Az informatikai iparral ellentétben, ahol a nyílt és harmonizált adatformátumok váltak szabványossá, a CAD-ágazatban (BIM) minden szoftver a saját formátumát használja, ami zárt ökoszisztemákat hoz létre, és mesterségesen korlátozza a felhasználókat.



FEJEZET 6.2.

ZÁRT PROJEKTFORMÁTUMOK ÉS INTEROPERABILITÁSI PROBLÉMÁK

Zárt adatok és csökkenő termelékenység: a CAD-ipar zsákutcája (BIM)

A CAD -rendszer tulajdonosi jellege azt eredményezte, hogy minden programnak saját, egyedi adatformátuma van, amely vagy zárt és kívülről nem hozzáférhető - RVT, PLN, DWG, NDW, NWD, SKP, vagy félgy strukturált formában, meglehetősen bonyolult konverziós folyamat révén áll rendelkezésre - JSON, XML (CPIXML), IFC, STEP és ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV stb. stb..

A különböző adatformátumok, amelyekben ugyanazok a projektek azonos adatai tárolhatók, nemcsak szerkezetükben különböznek, hanem a belső jelölés különböző verziót is tartalmazzák, amit a fejlesztőknek figyelembe kell venniük az alkalmazások kompatibilitásának biztosítása érdekében. Például egy 2025-ös CAD-formátum megnyílik egy 2026-os CAD-programban, de ugyanaz a projekt soha nem fog megnyílni a CAD-program minden olyan verziójában, amely esetleg 2025 előtt volt elérhető.

Azáltal, hogy nem biztosít közvetlen hozzáférést az adatbázisokhoz, az építőipari szoftverszolgáltató gyakran saját egyedi formátumot és eszközöket hoz létre, amelyeket a szakembernek (tervezőmérnöknek vagy adatkezelőnek) kell használnia az adatok eléréséhez, importálásához és exportálásához.

Ennek következtében az alapvető CAD (BIM) és a kapcsolódó megoldások (pl. ERP/PMIS)) szállítói folyamatosan emelik a termékek használatának árait, és az egyszerű felhasználók kénytelenek "jutalékot" fizetni a formátumok szerinti adatátvitel minden egyes szakaszában [63]: a csatlakoztatásért, importálásért, exportálásért és a felhasználók által saját maguk által létrehozott adatokkal való munkáért.

A népszerű CAD - (BIM-) termékek felhőalapú adattárolásban történő adatelérésének költsége 2025-re eléri az 1 dollárt tranzakciónként [120], az építőipari ERP -termékek előfizetései pedig a középvállalkozások számára évente öt- és hat számjegyű összegeket tesznek ki [121].

A modern építőipari szoftverek lényege, hogy nem az automatizálás vagy a hatékonyság növelése, hanem a mérnökök képessége egy adott, magasan specializált szoftver megértésére befolyásolja az építési projektek adatfeldolgozásának minőségét és költségeit, valamint az építési projekteket végző vállalatok nyereségét és hosszú távú túlélését.

Az adatbázisokhoz való hozzáférés hiánya CAD -rendszer, amelyeket több tucat más rendszerben és több száz folyamatban használnak [63], és az ebből következően az egyes szakemberek közötti minőségi kommunikáció hiánya az építőipart a gazdaság egyik legkevésbé hatékony ágazatává tette a termelékenység szempontjából [44].

A CAD- (BIM-) tervezési alkalmazások, az új rendszerek (ERP), az új építési technológiák és anyagok megjelenése során az elmúlt 20 évben a teljes építőipar termelékenysége 20%-kal csökkent (2.2-1. ábra), míg a gazdaság minden olyan ágazatának teljes termelékenysége, amely nem küzd jelentős problémákkal az adatbázisokhoz való hozzáféréssel és a marketingszerű BIM-koncepciókkal, 70%-kal nőtt (96%-kal a feldolgozóiparban) [122].



6.2-1. ábra A projektadatok elszigeteltsége és összetettsége miatt, amelytől az építőiparban több tucat osztály és több száz folyamat függ, a döntéshozatal sebessége többszörösen lassabb, mint más iparágakban.

Vannak azonban elszigetelt példák a CAD-megoldások közötti átváthatóság megteremtésének alternatív megközelítéseire is. Európa legnagyobb építőipari vállalata a még 2018-ban indult SCOPE projekttel [123] mutatja be, hogy hogyan lehet túllépni a CAD- (BIM-) rendszerek klasszikus logikáján. Ahelyett, hogy megpróbálnák alárendelni az IFC vagy szabadalmaztatott geometriamagokra támaszkodnának, a SCOPE fejlesztői API-kat és SDK-kat reverse engineering használnak, hogy különböző CAD programokból adatokat vonjanak ki, azokat semleges formátumokká, például OBJ vagy CPIXML formátumokba konvertálják az egyetlen nyílt forráskódú geometriamag OCCT alapján, és tovább alkalmazzák az építőipari és tervező cégek több száz üzleti folyamatában. Az ötlet progresszivitása ellenére azonban az ilyen projektek szembesülnek a szabad geometriai rendszermagok korlátaival és összetettségével, és továbbra is egy vállalat zárt ökoszisztémáinak részei maradnak, amelyek a monovendor megoldások logikáját reprodukálják.

A zárt rendszerek korlátai és az adatformátumok közötti különbségek, valamint az egységesítésükhez szükséges hatékony eszközök hiánya miatt a CAD-formátumokkal dolgozó vállalatoknak jelentős mennyiségi, eltérő szerkezetű és zártságú adathalmazzal kell szembenézniük. Ezeket az adatokat nem használják fel megfelelően, és eltűnnék az archívumokban, ahol örökre elfelejtve és kihasználatlanul maradnak.

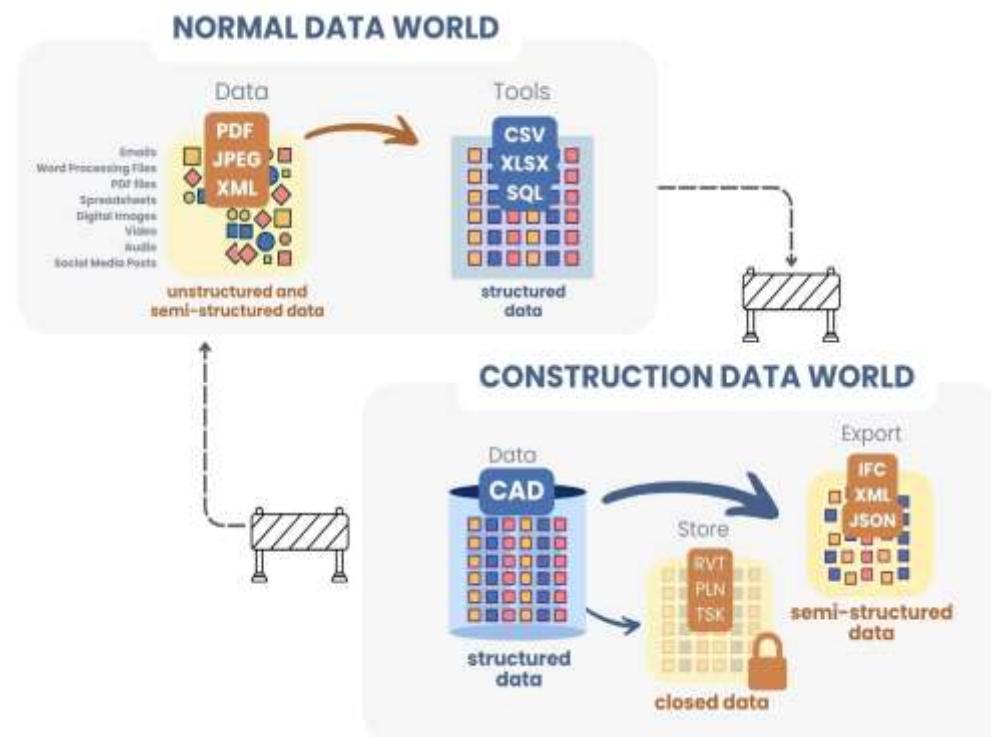
A tervezési fázisban jelentős erőfeszítésekkel nyert adatok összetettsége és zárt jellege miatt a további felhasználás szempontjából hozzáférhetetlenné válnak.

Ennek eredményeként az elmúlt 30 évben az építőipari fejlesztők újra és újra ugyanazzal a problémával szembesültek: minden egyes új zárt formátumú vagy szabadalmaztatott megoldás szükségessé teszi a meglévő nyílt és zárt CAD-rendszerekkel való integrációt. A különböző CAD- és BIM-megoldások közötti

átjárhatóság biztosítására tett folyamatos kísérletek csak bonyolítják az adatökoszisztémát, ahelyett, hogy hozzájárulnának annak egyszerűsítéséhez és szabványosításához.

Az interoperabilitás mítosza a CAD-rendszerek között

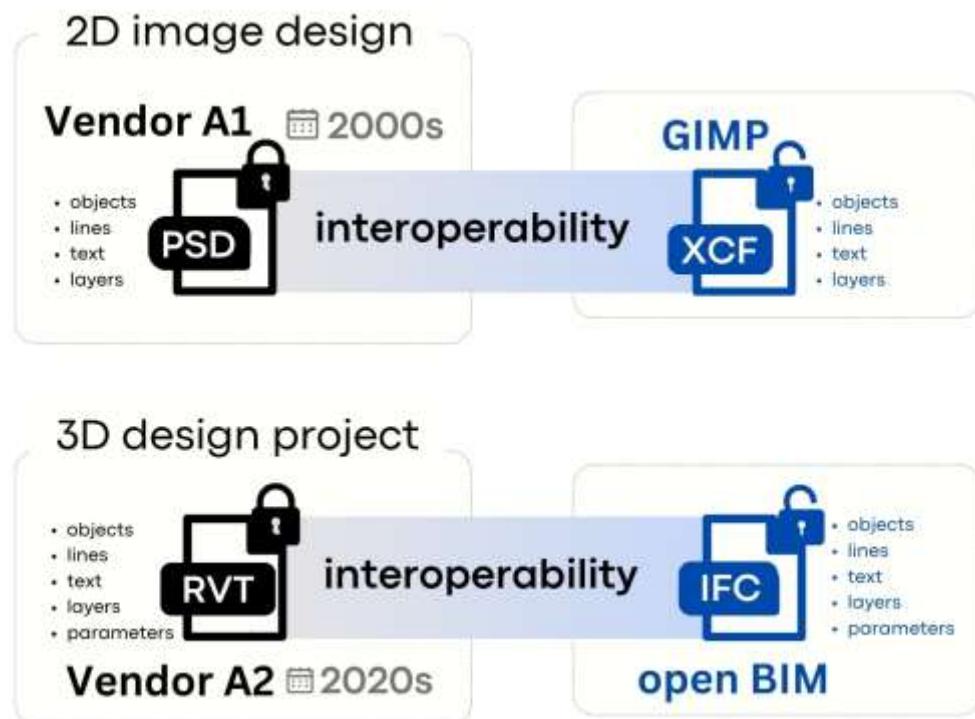
Míg az 1990-es évek közepén a CAD-környezetben az interoperabilitás fejlesztésének fő iranya a szabadalmaztatott DWG formátum megtörése volt - ami az Open DWG szövetség [75] győzelmeiben és az építőipar egész számára legnépszerűbb rajzformátum tényleges megnyitásában csúcosodott ki -, addig a 2020-as évek közepére a hangsúly eltolódott. Az építőiparban egy új trend kezd lendületet venni: számos fejlesztőcsapat arra összpontosít, hogy úgynevezett "hidakat" hozzon létre a zárt CAD-rendszerek (zárt BIM), az IFC formátum és a nyílt megoldások (nyílt BIM) között. E kezdeményezések többsége az IFC formátum és az OCCT geometria-mag használatán alapul, technikai hidat képezve az eltérő platformok között. Ezt a megközelítést ígéretes iránynak tekintik, amely jelentősen javíthatja az adatcserét és a szoftvereszközök interoperabilitását.



6.2-2. ábra Míg más iparágak nyílt adatokkal dolgoznak, az építőiparnak zárt vagy lazán strukturált CAD-formátumokkal kell dolgoznia (BIM).

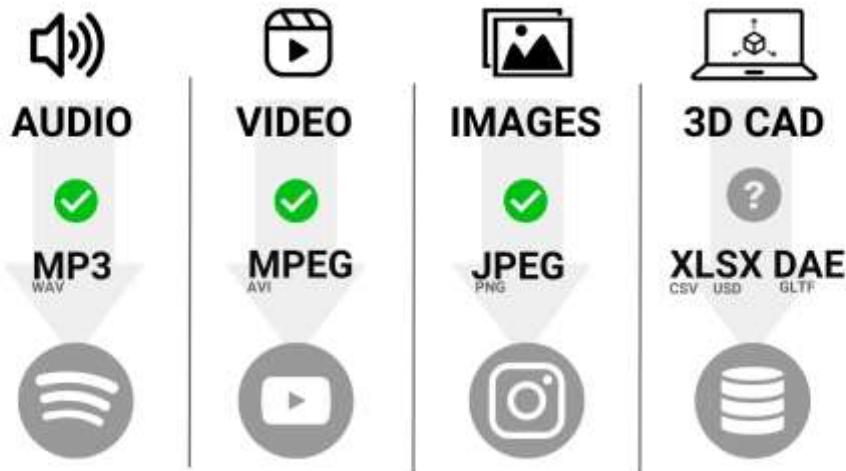
Ennek a megközelítésnek történelmi párhuzamai vannak. A 2000-es években a fejlesztők a legnagyobb grafikus szerkesztőprogram-gyártó (2D world) dominanciáját leküzdve megpróbáltak zökkenőmentes integrációt létrehozni saját fejlesztésű megoldása és az ingyenes nyílt forráskódú között - a GIMP alternatívájaként (6.2-3. ábra). Akkor is, mint ma az építésben, arról volt szó, hogy megpróbálták áthidalni a zárt és a nyílt rendszereket, miközben megőrizték a komplex paramétereket, rétegeket és a szoftver belső logikáját.

A felhasználók azonban valójában egyszerű megoldásokat kerestek - lapos, nyílt adatokat a rétegek és programparaméterek túlzott bonyolultsága nélkül (a CAD geometriai magjának analógjai). A felhasználók egyszerű és nyílt adatformátumokat kerestek, amelyek mentesek a túlzott logikától. A grafikában a JPEG, a PNG és a GIF ilyen formátumok lettek. Ma ezeket használják a közösségi hálózatokban, a weboldalakon, az alkalmazásokban - platformtól és szoftvergyártótól függetlenül könnyen feldolgozhatók és értelmezhetők.



6.2-3 ábra Az adatformátumok átjárhatósága az építőiparban hasonló ahhoz, amelyet a 2000-es években egy népszerű gyártó saját fejlesztésű termékének és a nyílt forráskódú GIMP-nek az egyesítésére tett kísérletek jártak be.

Ennek eredményeképpen a képalkotó iparban ma már szinte senki sem használ zárt formátumokat, például PSD-t vagy nyílt XCF-et alkalmazásokhoz, közösségi hálózatokhoz, mint a Facebook és az Instagram, vagy weboldalak tartalmaként. Ehelyett a legtöbb feladatra a sima és nyílt JPEG, PNG és GIF formátumokat használják az egyszerű használat és a széles körű kompatibilitás miatt. Az olyan nyílt formátumok, mint a JPEG és a PNG, sokoldalúságuk és széles körű támogatásuk miatt váltak a képmegosztás szabványává, így számos platformon könnyen használhatók. Hasonló átmenet figyelhető meg más csereformátumok, például a videó és az audio esetében is, ahol az olyan univerzális formátumok, mint az MPEG és az MP3, tömörítési hatékonyságuk és széles körű kompatibilitásuk miatt kerültek előtérbe. A szabványosítás felé való ilyen elmozdulás leegyszerűsítette a tartalmak és információk megosztását és lejátszását, így azok több platformon keresztül minden felhasználó számára elérhetővé váltak (6.2-4. ábra).



6.2-4 ábra Az adatok megosztására és felhasználására népszerűvé váltak a bonyolult szerkesztési funkciók nélküli, egyszerűsített formátumok.

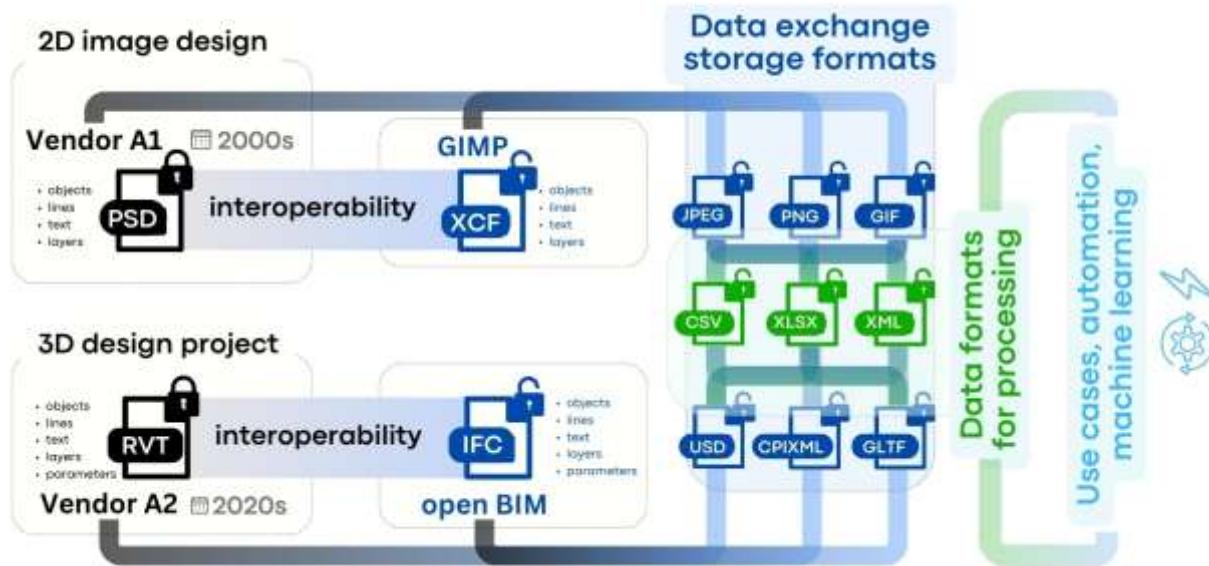
Hasonló folyamatok játszódnak le a 3D modellezés során. Az olyan egyszerű és nyílt formátumokat, mint az USD, OBJ, glTF, DAE, DXF, SQL és XLSX egyre gyakrabban használják a projektekben a CAD-környezeten kívüli adatcserére (BIM). Ezek a formátumok minden szükséges információt tárolnak, beleértve a geometriát és a metaadatokat, anélkül, hogy bonyolult BREP-struktúrát, geometriamagokat vagy gyártóspezifikus belső osztályozókat kellene működtetni. A vezető szoftvergyártók által biztosított saját formátumok, mint például az NWC, SVF, SVF2, CPIXML és CP2 szintén hasonló funkciókat látnak el, de a nyílt szabványokkal ellentétben zártak maradnak.

Figyelemre méltó (és érdemes újra felidézni, amint azt az előző fejezetben már említettük), hogy ezt az elképzelést - az olyan köztes semleges és parametrikus formátumok, mint az IGES, a STEP és az IFC elutasítását - már 2000-ben támogatta az a nagy CAD gyártó, amely a BIM Whitepaper című dokumentumot készítette és az IFC formátumot 1994-ben regisztrálta. A 2000. évi "Integrált tervezés és gyártás" című Whitepaperben [65] a CAD-szállító hangsúlyozza a CAD-adatbázishoz való natív hozzáférés fontosságát a szoftverkörnyezeten belül, anélkül, hogy közbenső fordítókat és parametrikus formátumokat kellene használni az információk teljességének és pontosságának megőrzése érdekében.

Az építőiparnak még nem sikerült megállapodnia sem a CAD-adatbázisokhoz való hozzáférés eszközeiről, sem azok kényszerű visszafejtéséről, sem a CAD-platformokon kívül használható közös egyszerűsített adatformátum elfogadásáról (BIM). Például Közép-Európában és a német nyelvű régiókban számos, az építőiparban működő nagyvállalat használja a CPIXML formátumot az ERP-rendszerében [121]. Ez a szabadalmaztatott formátum, amely egyfajta XML, a CAD (BIM) projektadatokat, beleértve a geometriai és metaadatokat is, egyetlen szervezett, egyszerűsített struktúrában egyesíti. A nagy építőipari vállalatok saját új formátumokat és rendszereket is létrehoznak, mint például a SCOPE projekt, amelyet az előző fejezetben tárgyaltunk

A parametrikus CAD-formátumok vagy az IFC (STEP) komplex parametrikus fájlok zárt logikája a legtöbb üzleti folyamatban felesleges. A felhasználók egyszerűsített és lapos formátumokat keresnek, mint például USD, CPIXML, XML &OBJ, DXF, glTF, SQLite, DAE &XLSX, amelyek tartalmazzák az összes szükséges

eleminformációt, de nem terhelik őket felesleges BREP-geometriai logikával, a geometriamagoktól való függőséggel és az egyes CAD és BIM -termékek belső osztályozásával (ábra). 6.2-5).



6.2-5 A legtöbb felhasználási esethez a felhasználók a lehető legegyszerűbb, a szállítói programuktól független formátumokat választják.

Az olyan lapos képformátumok, mint a JPEG, PNG és GIF megjelenése, amelyek mentesülnek a gyártók belső motorjainak felesleges logikájától, több ezer interoperábilis megoldás kifejlesztését ösztönözte a grafikafeldolgozásra és felhasználására. Ez a retusáló és szűrőeszközökkel kezdve az olyan közösségi hálózatokig, mint az Instagram, a Snapchat és a Canva, olyan alkalmazások megjelenéséhez vezetett, ahol ezeket az egyszerűsített adatokat egy adott szoftverfejlesztőhöz való kötöttség nélkül lehet felhasználni.

A tervezési CAD-formátumok szabványosítása és egyszerűsítése számos új, felhasználóbarát és független eszköz megjelenését fogja ösztönözni az építési projektekkal való munkához.

A zárt geometriai magokhoz kötött szállítói alkalmazások bonyolult logikájától az egyszerűsített elemek könyvtárain alapuló univerzális, nyílt formátumok felé való elmozdulás megteremti a rugalmasabb, átláthatóbb és hatékonyabb adatkezelés feltételeit. Ez egyben az építési folyamatban részt vevő valamennyi fél számára - a tervezőktől a megrendelőkig és a karbantartó szolgálatokig - megnyitja az információkhöz való hozzáférést.

Mindazonáltal nagyon valószínű, hogy az előkvetkező években a CAD gyártók megróbálják majd újra elmozdítani a interoperabilitásról és a CAD-adatbázisokhoz való hozzáférésről szóló vitát. Ez már az "új" koncepciókról fog szólni - mint például a granuláris adatok, intelligens gráfok, "föderált modellek", digitális ikrek a felhőalapú adattárakban -, valamint olyan iparági szövetségek és szabványok létrehozásáról, amelyek folytatják a BIM és a nyílt BIM újtát. A vonzó terminológia ellenére az ilyen kezdeményezések ismét olyan eszközökkel válhatnak, amelyek a felhasználókat a saját ökoszisztemákon belül tartják. Erre példa az USD (Universal Scene Description) formátum aktív népszerűsítése, amely 2023-tól a platformok

közötti CAD (BIM) együttműködés "új szabványa" lesz.

Tovább az USD-hez és a szemcsés adatokhoz m

Az AOUSD szövetség [124] 2023-as létrejötte fontos fordulatot jelent az építőiparban. Egy új valóság kezdetének vagyunk tanúi, amelyet a CAD-gyártók alakítanak ki az építési adatok kezelésében, több jelentős változásnak keresztül. Az első jelentős változás a CAD -adatok megítélését érinti. A koncepcionális tervezés korai szakaszában részt vevő szakemberek egyre inkább felismerik, hogy a CAD-környezetben történő tervkészítés csak a kiindulópont. A tervezési folyamat során keletkező adatok végül az objektumok elemzésének, működtetésének és kezelésének alapjává válnak. Ez azt jelenti, hogy ezeknek a hagyományos CAD-eszközökön túlmutató rendszerekben is hozzáférhetőnek és használhatónak kell lenniük.

Ezzel párhuzamosan forradalom zajlik a vezető fejlesztők megközelítésében. Az iparág vezető CAD-szállítója, a BIM koncepció és az IFC formátum megalkotója váratlan fordulatot vesz stratégiájában. A vállalat 2023-tól eltávolodik az adatok hagyományos, egyedi fájlokban történő tárolásától, a granuláris (normalizált és strukturált) adatokkal való munkára összpontosít, és adatközpontú megközelítésre tér át [125].

A gyártók más iparágak történelmi trendjeit követik: a legtöbb felhasználónak nincs szüksége zárt CAD-formátumokra (hasonlóan a PSD-hez) vagy összetett parametrikus IFC-fájlokra (hasonlóan a GIMP-hez, réteglogikával). Egyszerű objektumképekre van szükségük, amelyek felhasználhatók a CAFM (építőipari Instagram), az ERP (Facebook) és ezernyi más, Excel táblázatokkal és PDF dokumentumokkal teli folyamatban.

Az építőiparban jelenleg tapasztalható tendenciák potenciálisan előkészítik a terepet a parametrikus és összetett formátumoktól való fokozatos eltávolodásra az univerzálisabb és függetlenebb USD, GLTF, DAE, OBJ (a metainformációkkal minden a hibridben, minden pedig külön strukturált vagy lazán strukturált formátumokban) javára. A történelmi vezetők, közük a nagy tervezőcégek, amelyek az 1990-es évek közepén egykor aktívan támogatták az IFC-t, ma már nyíltan az új USD formátumot [93] népszerűsítik, hangsúlyozva annak egyszerűségét és sokoldalúságát (6.2-6. ábra). Az USD tömeges elterjedése a termékekben, a GLTF-kompatibilitás és az olyan eszközökbe való aktív integráció, mint a Blender, az Unreal Engine és az Omniverse, az adatokkal való munka új paradigmájának kezdetének lehetőségét mutatja. Az olyan lokalizált megoldások népszerűsége mellett, mint az európai lapos USD formátum - CPIXML, amelyet a népszerű európai ERP-kben használnak, az USD pozíciója megerősödhet Közép-Európában. Az IFC formátum fejlesztésében részt vevő szervezetek már most az USD-hez igazítják stratégiájukat [126], ami csak megerősíti a váltás elkerülhetetlenségét.

Technical Specifications				Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility	
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data	
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.	
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON	
Export/import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products	
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it.	
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse	
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms	
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools	
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions	
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications	
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure	
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations	
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement	

6.2-6. ábra Az IFC és az USD formátum technikai specifikációinak összehasonlítása.

Mindezek fényében az USD potenciálisan de facto szabvánnyá válhat, és ígéretesnek tűnik számos jelenlegi korlátot leküzdjeni, amelyek elsősorban a meglévő CAD - (BIM-) formátumok összetettségével és értelmezésüknek a geometriai maguktól való függőségével kapcsolatosak.

A parametrikus és összetett CAD -formátumok és az IFC helyett az egyszerűsített adatformátumok USD, gLTF, DAE, OBJ, az elemek metainformációival együtt CSV, XLSX, JSON, XML -ban fognak helyet kapni az építőiparban, egyszerűségük és rugalmasságuk miatt.

Az építőipar jelenlegi változásai első pillantásra technológiai áttörésnek tűnnek, amely az elavult IFC -ról a modernebb USD -re való áttéréssel kapcsolatos. Érdemes azonban figyelembe venni, hogy még 2000-ben

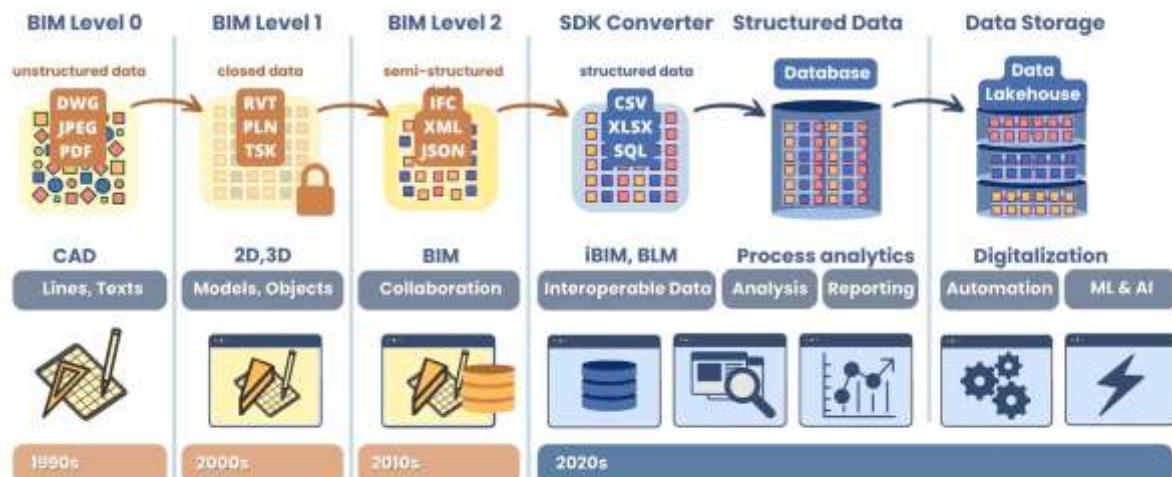
ugyanaz a CAD gyártó, amely az IFC-t kifejlesztette, írt annak problémáiról és az adatbázishoz való hozzáférés szükségességről [65], és most aktívan támogatja az új szabványra - USD - való áttérést.

A "nyílt adatok" USD és a granuláris adatkezelés "új" koncepciói mögött a felhőalapú alkalmazásokon keresztül, amelyeket a CAD gyártók kezdenek népszerűsíteni, a gyártóknak a projektadatok kezelésének monopolizálására irányuló szándéka rejtőzhet, ahol a felhasználók olyan helyzetben találják magukat, hogy a formátum kiválasztása inkább a vállalati érdekekhez, mint a valós igényekhez kapcsolódik.

A legfontosabb tények elemzése [93] azt mutatja, hogy ezeknek a változásoknak a fő célja kevésbé a felhasználói kényelem, hanem elsősorban az ökoszisztemák és az adatáramlás feletti ellenőrzés fenntartása a gyártók javára, akik 40 év alatt soha nem tudtak hozzáférést biztosítani a CAD-adatbázisokhoz.

Talán itt az ideje, hogy a vállalatok ne várjanak tovább a szoftvergyártók új koncepcióira, hanem az adatközpontú irányba történő saját fejlesztésre összpontosítanak. Miután a reverse engineering eszközökkel megszabadult az adathozzáférési problémáktól, az iparág képes lesz önállóan, új koncepció erőltetése nélkül is átérni az adatok feldolgozására és elemzésére szolgáló modern, ingyenes és kényelmes eszközökre.

CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



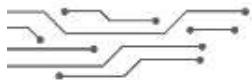
Ábra 6.2-7 CAD érettségi szint (BIM): a strukturálatlan adatoktól a strukturált adatokig és a tárolóig.

Az adatbázisokhoz való hozzáférés, a nyílt adatok és formátumok elkerülhetetlenül standarddá válnak az építőiparban, függetlenül attól, hogy a gyártók megpróbálják-e feltartóztatni a folyamatot - ez csak idő kérdése (6.2-7. ábra). Az átmenet üteme jelentősen felgyorsulhat, ha egyre több szakember ismerkedik meg a nyílt formátumokkal, az adatbázis-eszközökkel és az elérhető reverse engineering SDK-kkel, amelyek lehetővé teszik a közvetlen hozzáférést a CAD-adatokhoz -rendszerük [92].

A jövő a nyílt, egységes és analitikusan hozzáférhető adatokban rejlik. A szállítói megoldásoktól való függés elkerülése és a zárt ökoszisztemák túszaiivá válás elkerülése érdekében az építőipari és mérnöki vállalatoknak előbb-utóbb a nyitottságra és a függetlenségre kell támaszkodniuk, és olyan formátumokat és megoldásokat kell választaniuk, amelyek teljes ellenőrzést biztosítanak az adatok felett.

Az építőiparban ma keletkező adatok a jövőben az üzleti döntések kulcsfontosságú forrásai lesznek. Stratégiai "üzemanyagként" fog működni, amely az építőipari vállalatok fejlődését és hatékonyságát táplálja. Az építőipar jövője az adatokkal való munkavégzés képességében rejlik, nem pedig az adatformátumok vagy modellek kiválasztásában.

Az USD, glTF, DAE, OBJ és a szabadalmaztatott parametrikus CAD formátumok közötti különbség megértéséhez fontos megvizsgálni a vizualizáció és a tervezési számítások egyik legösszetettebb és legfontosabb adatelemét - a geometriát és annak generálási folyamatait. Annak megértéséhez pedig, hogy a geometriai adatok hogyan válnak az építőipari analitika és számítások alapjává, mélyebben be kell tekinteni a geometria létrehozásának, átalakításának és tárolásának mechanizmusába.



FEJEZET 6.3.

GEOMETRIA AZ ÉPÍTŐIPARBAN: A VONALAKTÓL A KÖBMÉTEREKIG

Amikor a vonalak pénzzé válnak, vagy hogy miért van szükségük az építőknek geometriára

A geometria az építőiparban nemcsak a szemléltetés, hanem a pontos mennyiségi számítások alapja is. A projektmodellben a geometria kiegészíti az elemparaméterek listáit (3.1-16. ábra) olyan fontos térfogati jellemzőkkel, mint a hossz, a terület és a térfogat. Ezek a térfogati paraméterétek automatikusan kiszámításra kerülnek a geometria magok segítségével és a becslések, ütemtervek és erőforrásmodellek kiindulópontját képezik. Amint azt már e könyv 5. részében és a "Költségszámítás és becslés építési projektek" című fejezetben tárgyaltuk, a CAD -modellekből származó objektumcsoporthok térfogati paraméterei képezik a modern ERP, PMIS -rendszerök alapját/ A geometria nemcsak a tervezési fázisban játszik alapvető szerepet, hanem a projekt megvalósításának irányításában, az ütemezés ellenőrzésében, a költségvetés készítésében és az üzemeltetésben is. Ahogy évezredekkel ezelőtt, az egyiptomi piramisok építésekor a projekt pontossága olyan hosszmértékektől függött, mint a cubit és a cubit, úgy ma a CAD -programokban a geometria értelmezésének pontossága közvetlenül befolyásolja az eredményt: a költségvetéstől és a határidőtől kezdve a vállalkozó kiválasztásán át a szállítási logisztikáig

A rendkívül versenyképes és költségvetési szempontból korlátozott környezetben a térfogatszámítások pontossága, amely közvetlenül függ a geometriától, túlélési tényezővé válik. A modern ERP -rendszerök közvetlenül függnek a CAD - és BIM -modellekből nyert helyes térfogati jellemzőktől. Ezért az elemek pontos geometriai leírása nem csupán vizualizáció, hanem a költségek és az építési idő kezelésének kulcsfontosságú eszköze.

Történelmileg a mérnöki kommunikáció elsődleges nyelve a geometria volt. A papiruszon lévő vonalaktól a digitális modellekig a rajzok és geometriai ábrázolások a tervezők, művezetők és becslők közötti információcsere eszközöként szolgáltak. A számítógépek megjelenése előtt a számításokat kézzel, vonalzó és szögmérő segítségével végezték. Ma ez a feladat a térfogatmodellezésnek köszönhetően automatizált: a CAD-szoftverek geometriai magjai a vonalakat és pontokat háromdimenziós testekké alakítják, amelyekből automatikusan kivonják az összes szükséges jellemzőt.

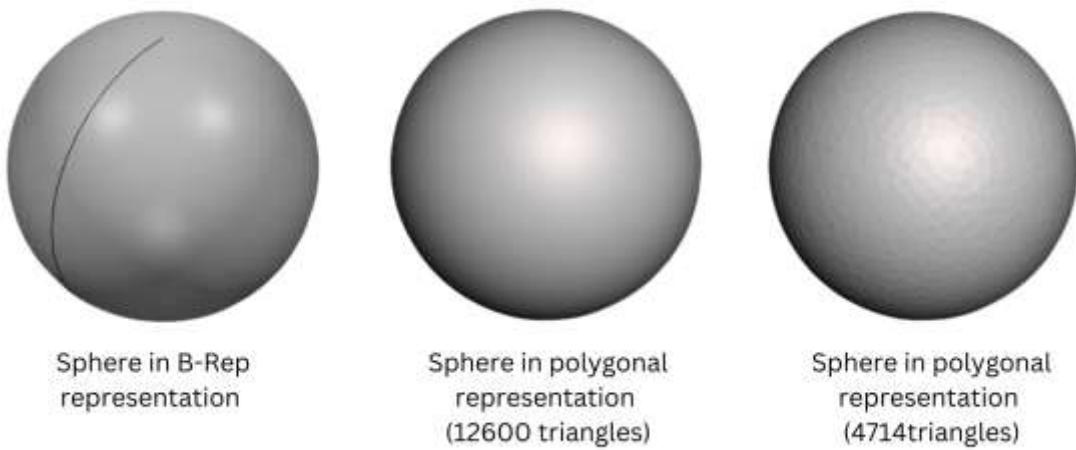
A CAD- (BIM-) programok felhasználói felületén keresztül történik a számításokhoz szükséges geometriai elemek létrehozása a CAD- (BIM-) -programokban. A pontok és vonalak térfogati testekké való átalakításához a geometriai magot használják, amely a legfontosabb feladatot - a geometria térfogati modellekkel való átalakítását - látja el, amelyből a közelítés után automatikusan kiszámításra kerülnek az elem térfogati jellemzői.

Vonalakból kötetekké: Hogyan lesz a terület és a térfogat adat

A mérnöki gyakorlatban a térfogatokat és területeket geometriai felületekből számítják ki analitikusan vagy parametrikus modellekkel, mint például a NURBS (nonuniform rational B-splines) a BREP (boundary element representation) keretrendszerben.

A NURBS (Non-Uniform Rational B-Splines) a görbék és felületek leírásának matematikai módja, míg a BREP egy olyan keretrendszer, amely egy tárgy teljes háromdimenziós geometriájának leírására szolgál, beleértve a NURBS segítségével definiálható határokat is.

A BREP és a NURBS pontossága ellenére nagy számítási erőforrásokat és összetett algoritmusokat igényelnek. Az ilyen matematikailag pontos leírásokból történő közvetlen számítás azonban gyakran számítási nehézségekbe ütközik, ezért a gyakorlatban szinte minden a tesszellációt - a felületek háromszögek rácsává alakítását - alkalmazzák, ami egyszerűsíti a későbbi számításokat. A tesszelláció egy összetett felület háromszögekre vagy sokszögekre való felosztása. A CAD /CAE környezetekben ezt a módszert használják a vizualizációhoz, a térfogatszámításhoz, az ütközési kereséshez, az olyan formátumokba való exportáláshoz, mint a MESH és az ütközéselemzéshez. A természetből vett példa a méhsejt, ahol egy összetett alakzatot szabályos rácsra bontanak (6.3-1. ábra).



Ábra 6.3-1 Ugyanaz a gömb a BREP parametrikus leírásában és a háromszögek különböző számú poligonális .

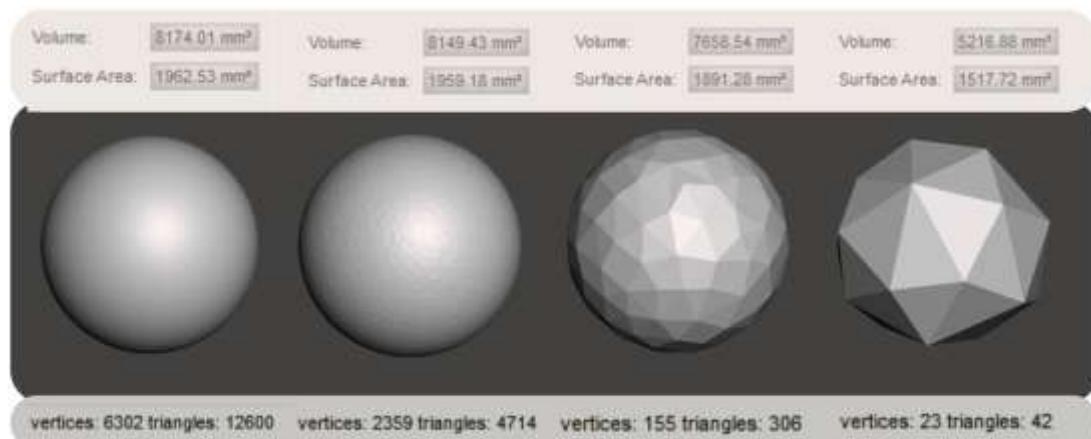
A BREP (NURBS), amelyet a CAD -ben használnak, nem alapvető geometriai modell. Kényelmes eszközöként jött létre a körök és racionális spline-ok ábrázolásához, valamint a geometriai adatok tárolásának minimalizálásához. Vannak azonban korlátai - például az, hogy nem képes pontosan leírni a spirális vonalak és felületek alapjául szolgáló szinuszoidot, és hogy bonyolult geometriai magokat kell használni.

Ezzel szemben a háromszöghálót és a parametrikus alakzatok tesszellálását az egyszerűség, a hatékony memóriahasználat és a nagy mennyiségű adat feldolgozásának képessége jellemzi (6.3-2. ábra). Ezek az előnyök lehetővé teszik, hogy a geometriai alakzatok számításakor lemondunk a bonyolult és drága geometriai kernelekről, és az ezekbe ágyazott több tízmillió sornyi kódrról.

A legtöbb épület esetében nem számít, hogy a térfogati jellemzőket pontosan hogyan határozzák meg - parametrikus modellekben (BREP, IFC) vagy poligonokon keresztül (USD, glTF, DAE, OBJ). A geometria továbbra is a közelítés formája marad: akár NURBS, akár MESH révén, mindig az alakzat közelítő leírásáról van szó.

A poligonokként vagy BREP (NURBS) meghatározott geometria bizonyos mértékig csak egy folytonos alakzat közelítő leírásának módja marad. Ahogyan a Fresnel-integráloknak sincs pontos analitikus kifejezése, a geometria poligonok vagy NURBS segítségével történő diszkretizálása mindig közelítés, ahogyan a háromszög MESH is.

A BREP formátumú parametrikus geometriára elsősorban ott van szükség, ahol fontos a minimális adatmérét, és a feldolgozáshoz és megjelenítéshez erőforrás-igényes és drága geometriamagokat lehet használni. Leggyakrabban ez a CAD -programok fejlesztőire jellemző, akik erre a célra az MCAD -szolgáltatók geometriai magjait alkalmazzák termékeikben. Ebben az esetben a BREP-modelleket a tesszellálás során a megjelenítéshez és a számításokhoz még ezekben a programokban is gyakran háromszögekké alakítják át (hasonlóan ahhoz, ahogyan a PSD-fájlokat JPEG-re egyszerűsítik).



6.3-2. ábra A térfogati jellemzők különbsége a különböző sokszögszámú ábrákon.

A poligonális MESH, valamint a parametrikus BREP, saját előnyökkel és korlátokkal rendelkezik, de a cél ugyanaz - a geometria leírása a felhasználó feladatainak figyelembevételével. Végső soron egy geometriai modell pontossága nemcsak az ábrázolás módszerétől, hanem az adott feladat követelményeitől is függ.

A legtöbb konstrukciós probléma esetén a parametrikus geometria és a bonyolult geometriai magok szükségessége felesleges lehet.

Minden egyes számításautomatizálási feladat esetében érdemes megfontolni, hogy a CAD-fejlesztők, akik saját szoftvertermékeik népszerűsítésében és értékesítésében érdekeltek, nem túlozzák-e el a parametrikus geometria jelentőségét.

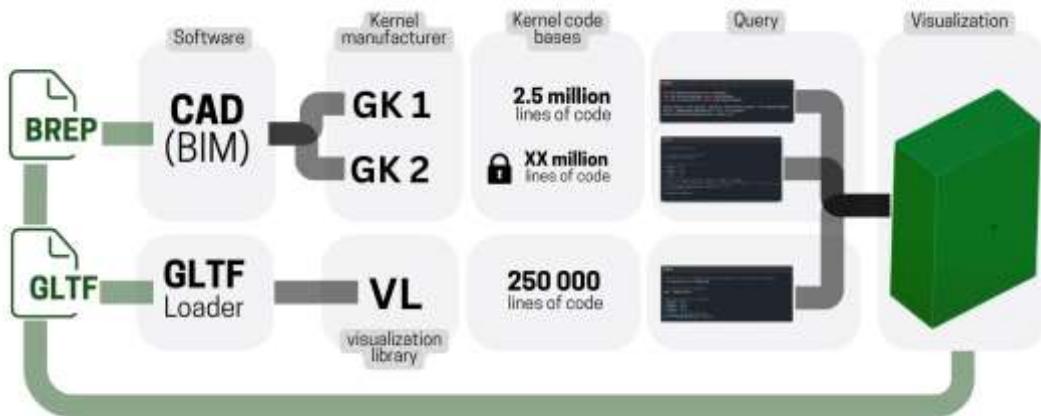
Mozgás a MESH, USD és poligonok felé: a tesszelláció használata a geometriához

Az építőiparban a tervezési információkkal és a jellemző geometriával dolgozó rendszerek, adatbázisok áramoltatása, fejlesztése vagy folyamatok automatizálása során fontos, hogy törekedjünk a konkrét CAD-szerkesztőprogramuktól és geometriamaguktól való függetlenségre.

A számítási részlegekben és az építkezésen egyaránt használandó csereformátumnak nem szabad egy adott CAD- (BIM-) programon alapulnia. A geometriai információkat a formátumban közvetlenül tesszellációval kell ábrázolni, a geometriai magra vagy a CAD-architektúrára való hivatkozás nélkül.

A CAD oldalról származó parametrikus geometria köztes forrásnak tekinthető, de nem egy univerzális formátum alapjául. A legtöbb parametrikus leírást (beleértve a BREP és NURBS) a további feldolgozáshoz mindenkorban poligonális MESH -re kell konvertálni. Ha az eredmény ugyanaz (tesszelláció és poligonok) és a folyamat egyszerűbb, a választás kézenfekvő. Ez analóg a gráfontológiák és a strukturált táblázatok közötti választással (amelyet a negyedik részben tárgyalunk): a túlzott bonyolultság ritkán indokolt (3.2-10. ábra, 6.1-8. ábra).

Az olyan nyílt formátumok, mint: OBJ, STL, glTF, SVF, CPIXML, USD és DAE, univerzális háromszögletes hálószerkezetet használnak, ami jelentős előnyökkel jár. Ezek a formátumok kiváló interoperabilitással rendelkeznek - könnyen olvashatók és vizualizálhatók a rendelkezésre álló nyílt forráskódú könyvtárak segítségével anélkül, hogy bonyolult, több millió sornyi kódot tartalmazó speciális geometriai magokra lenne szükség (6.3-3. ábra). Ezeket a sokoldalú geometriaformátumokat az IKEA™ viszonylag egyszerű konyhai tervezőszközöktől kezdve a mozi és a VR -alkalmazásokban használt komplex objektumvizualizációs rendszerekig sokféle alkalmazásban használják. Fontos előnye, hogy számos ingyenes és nyílt forráskódú könyvtár áll rendelkezésre az ilyen formátumokkal való munkához, amelyek a legtöbb platformra és programozási nyelvre elérhetők.



6.3-3. ábra Ugyanez a geometria ábrázolás érhető el parametrikus formátumok és geometriamagok, illetve háromszögelt formátumok és nyílt forráskódú vizualizációs könyvtárak használatával.

A felhasználókon kívül a CAD -szolgáltatóknak is gondot okoz a külföldi parametrikus CAD-formátumok vagy a nyílt IFC értelmezése az eltérő geometriai magok miatt. A gyakorlatban kivétel nélkül minden CAD -szolgáltató a reverse engineering SDK -SDK-t használja a rendszerek közötti adatátvitelre, és egyikük sem támaszkodik olyan formátumokra, mint az IFC vagy az USD [93] az interoperabilitás érdekében.

Ahelyett, hogy a CAD-gyártók szövetségei által támogatott koncepciókat használnának, amelyeket ők maguk nem használnak - a CAD-megoldások fejlesztői és felhasználói számára produktívabb, ha az egyes megközelítések előnyeinek megértésére összpontosítanak egy adott kontextusban, és a felhasználási esettől függően választják ki az egyik vagy másik típusú geometriát. A különböző geometriai ábrázolások közötti választás kompromisszumot jelent a pontosság, a számítási hatékonyság és az adott feladat gyakorlati igényei között.

A geometriai magok használatához kapcsolódó bonyolultság, amelyet a nagy gyártók hagyományosan az építőiparra kényszerítének a tervezési adatok feldolgozása során, gyakran feleslegesnek bizonyul. A MESH geometrián alapuló USD formátum egyfajta "Pandora szelencéjévé" válhat az iparág számára, új lehetőségeket nyitva a fejlesztők számára az adatcseré megszervezésére - a CAD gyártókra jellemző IFC és parametrikus BREP struktúrák keretein kívül.

Az USD, DAE, gLTF, OBJ stb. struktúrájának közelebbi vizsgálata után nyilvánvalóvá válik, hogy léteznek egyszerűbb, nyílt formátumok, amelyek lehetővé teszik a geometriai információk átvitelének és felhasználásának hatékony megszervezését anélkül, hogy bonyolult parametrikára és zárt geometriai kernelekre kellene támaszkodni. Ez a megközelítés nemcsak a fejlesztők számára csökkenti a technikai belépési küszöböt, hanem a rugalmas, skálázható és valóban nyílt megoldások kifejlesztésének is kedvez a digitális építés számára.

LOD, LOI, LOMD - a részletek egyedi osztályozása a CAD-ben (BIM).

A geometriai ábrázolási formátumok mellett egy olyan világban, ahol a különböző iparágak különböző részletességű és mélységű adatokat használnak, a CAD - (BIM-) módszerek saját egyedi osztályozási rendszereket kínálnak, amelyek strukturálják az épületmodellek informálásának megközelítését.

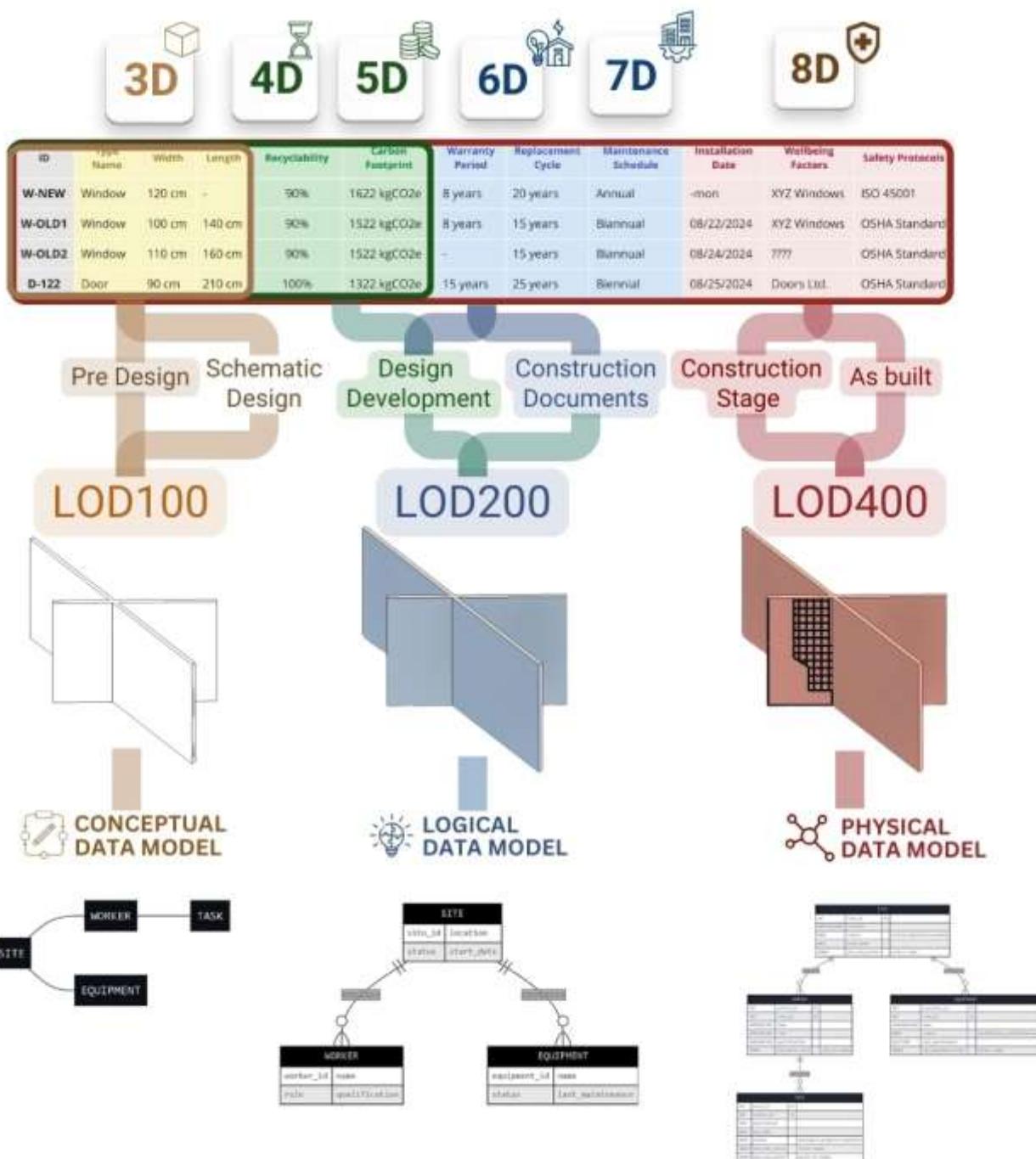
A szabványosítás új megközelítéseinek egyik példája a modellfejlesztési szintek bevezetése, amelyek a grafikus és információs komponensek készültségi fokát és megbízhatóságát tükrözik. A CAD - (BIM-) adatokkal végzett munka során az információtartalom megkülönböztetésére megjelent a LOD (Level Of Detail) - a modell grafikus részének részletességi szintje, és a LOI (Level Of Information) - az adatok kidolgozottsági szintje. Ezen kívül az integrált megközelítéshez bevezették a LOA (Level of Accuracy) - a bemutatott elemek pontossága és a LOG (Level of Geometry) fogalmát a grafikus ábrázolás pontosságának meghatározására.

A részletességi szinteket (LOD) 100-tól 500-ig terjedő számokkal jelölik, amelyek a modell fejlettségi fokát tükrözik. A LOD 100 egy koncepcionális modell általános formákkal és méretekkel. A LOD 200 pontosabb méreteket és formákat tartalmaz, de feltételes részletességgel. A LOD 300 egy részletes modell, pontos méretekkel, alakzatokkal és elemek helyével. A LOD 400 az elemek gyártásához és beépítéséhez szükséges részletes információkat tartalmazza. A LOD 500 a létesítmény építés utáni tényleges állapotát tükrözi, és az üzemeltetéshez és karbantartáshoz használatos. Ezek a szintek a CAD (BIM) modellinformációk telítettségének szerkezetét írják le az életciklus különböző szakaszaiban, beleértve a 3D, 4D, 5D és azon túl.

A valós projektekben a magas részletesség (LOD400) gyakran túlzott mértékű, és elegendő a LOD100 geometria vagy akár sík rajzok használata, míg a többi adatot vagy számítással, vagy olyan kapcsolódó elemekből lehet kinyerni, amelyek nem rendelkeznek különálló geometriával. Például a terek és a helyiségelemek (helyiségelem-kategóriák) nem rendelkeznek vizuális geometriával, de jelentős mennyiségű információt és adatbázist tartalmaznak, amelyek köré számos üzleti folyamat épül.

Ezért fontos, hogy a tervezés megkezdése előtt egyértelműen meghatározzuk a szükséges részletesség szintjét. A 4D -7D felhasználási esetek esetében gyakran még a DWG rajzok és a minimális LOD100 geometria is elegendő. A követelményekkel kapcsolatos folyamat során a legfontosabb feladat a modell gazdagsága és gyakorlatiassága közötti egyensúly megtalálása.

Lényegében, ha a CAD (BIM) adatokat adatbázisnak tekintjük (márpedig az), a modellel való telítettség új rövidítésekkel történő leírása nem más, mint az információs rendszerek lépésről lépésre történő adatmodellezése, a koncepcionális szinttől kezdve a fizikai szintig (6.3-4. ábra), amelyet a könyv harmadik és negyedik részében részletesen tárgyaltunk. A LOD és LOI minden egyes növelése új feladatokhoz szükséges információk hozzáadását jelenti: számítások, építésirányítás, üzemeltetés, és a modellnek további információs rétegekkel (3D -8D) való egymást követő gazdagítása jellemzi különböző paraméterek formájában, amelyeket a könyv ötödik részében tárgyaltunk.



6.3-4. ábra A projekt részleteinek konkretizálása megegyezik az adatmodellezéssel a koncepcionális adatmodelltől a fizikai adatmodellig.

A geometria csak egy része a tervezési adatoknak, amelyek szükségessége nem minden indokolt az építési projektekben, és a CAD -adatokkal való munka kulcskérdése nem annyira a modellek megjelenítésének módja, hanem inkább az, hogy a modellekből származó adatok hogyan használhatók fel a CAD- (BIM-) programokon kívül.

A 2000-es évek közepére az építőipar példátlan kihívással nézett szembe, mivel az irányítási és

adatfeldolgozó rendszerekben - különösen a CAD-részlegekből származó adatokban - gyorsan nőtt az adatmennyiség (BIM). Az adatmennyiség e drámai növekedése váratlanul érte a vállalatvezetőket, és nem voltak felkészülve az adatminőséggel és -kezeléssel szemben támasztott növekvő követelményekre.

Új CAD-szabványok (BIM) - AIA, BEP, IDS, LOD, COBie

Kihasználva a CAD-adatbázisokhoz való nyílt hozzáférés hiányát és a korlátozott versenyt az adatfeldolgozási piacon, valamint a BIM új rövidítéssel kapcsolatos marketingkampányokat felhasználva a CAD-adatok kezelésével kapcsolatos megközelítések kidolgozásában részt vevő szervezetek - új szabványok és koncepciók létrehozásába kezdtek, amelyeknek de jure az adatkezelési gyakorlatok javítását kell célozniuk.

Bár a CAD-gyártók és a fejlesztők (BIM) által közvetlenül vagy közvetve támogatott szinte valamennyi kezdeményezés a munkafolyamatok optimalizálását célozta, a különböző érdekeltek által lobbizott szabványok sokaságát eredményezte, ami az építőiparban az adatfolyamatok tekintetében némi félreérthetőséget és zavart eredményezett.

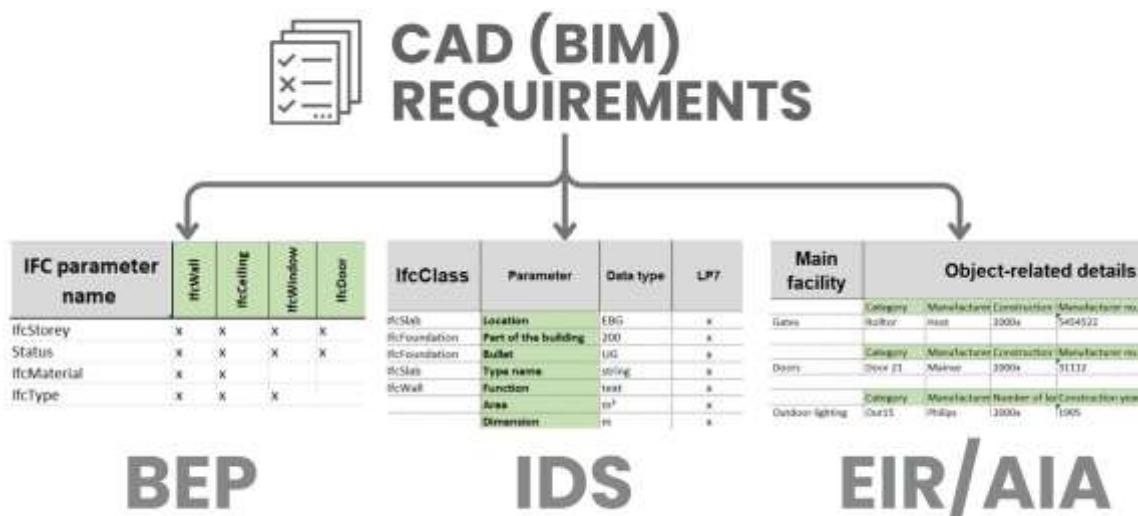
Soroljunk fel néhány új adatszabványt a LOD, LOI, LOA, LOG mellett, amelyek az elmúlt években jelentek meg az építőiparban:

- **BEP** (BIM Végrehajtási terv) - leírja, hogyan lehet a CAD (BIM) integrálni és felhasználni egy projektben, meghatározva az adatkezelési módszereket és folyamatokat.
- **EIR dokumentum /AIA** (az ajánlatkérő információs követelményei) - az ajánlatkérő által az ajánlattételi felhívás előtt készített dokumentum, amely tartalmazza a követelményeket a vállalkozóval szemben az információk elkészítésére és szolgáltatására vonatkozóan. Az adott projektben a BEP alapjául szolgál.
- **AIM** (Asset Information Model) a BIM folyamat része. A projekt átadása és befejezése után az adatmodellt eszközinformációs modellnek vagy AIM-nek nevezik. Az AIM célja a megvalósult eszköz kezelése, karbantartása és üzemeltetése.
- **IDS** (Information Delivery Specification) - meghatározza a követelményeit, valamint azt, hogy milyen adatokra és milyen formátumban van szükség az építési projekt különböző szakaszaiban.
- **iLOD** az a részletességi szint LOD, amellyel az információ a BIM -modellben reprezentálva van. Meghatározza, hogy a modellben szereplő információk mennyire részletesek és teljes körűek, az alapvető geometriai ábrázolásoktól a részletes specifikációig és adatokig.
- **eLOD** - LOD a CAD-modell egyes elemeinek részletességi szintje (BIM). Meghatározza az egyes elemek modellezésének mértékét és a kapcsolódó információkat, mint például a méretek, anyagok, teljesítményjellemzők és egyéb releváns attribútumok.
- **APS** (Platform Services) és a főbb CAD-szolgáltatók egyéb termékei (BIM) - leírja a kapcsolódó és nyílt adatmodellek létrehozásához szükséges eszközöket és infrastruktúrát.

Bár a CAD (BIM) szabványok - mint például a LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - bevezetésének deklarált célja az adatkezelés minőségének javítása és az automatizálási lehetőségek bővítése, a gyakorlatban használatuk gyakran túlzott bonyolultsághoz és a folyamatok

széttöredezetségéhez vezet. Ha a CAD (BIM) modellt egyfajta adatbázisnak tekintjük, nyilvánvalóvá válik, hogy e szabványok közül sok megismétli a más iparágakban az információs rendszerekkel való munka során régóta alkalmazott és hatékony megközelítéseket. Az egyszerűsítés és egységesítés helyett az ilyen kezdeményezések gyakran további terminológiai terhet jelentenek, és akadályozzák a valóban nyitott és rugalmas megoldások megvalósítását.

Figyelemre méltó, hogy ezek közül az új koncepciók közül sok valójában a könyv első részeiben részletesen tárgyalt modellezési és adatérényesítési folyamatok helyébe lép, amelyeket a gazdaság más ágazataiban már régóta alkalmaznak. Az építőiparban viszont a szabványosítási folyamat gyakran az ellenkező irányba halad - új adatleírási formátumok, új szabványok és új adatérényesítési koncepciók jönnek létre, amelyek nem minden vezetnek valódi egységességhöz és gyakorlati alkalmazhatósághoz. Ennek eredményeképpen a feldolgozás egyszerűsítése és automatizálása helyett az ágazat a szabályozás és a bürokrácia további szintjeivel szembesül (6.3-1. ábra), ami nem minden segíti a hatékonyseg növelését.



6.3-1 Az adat- és információtartalomra vonatkozó követelmények az attribútumok és határértékeik leírására korlátozódnak, amelyeket táblázatokkal írnak le.

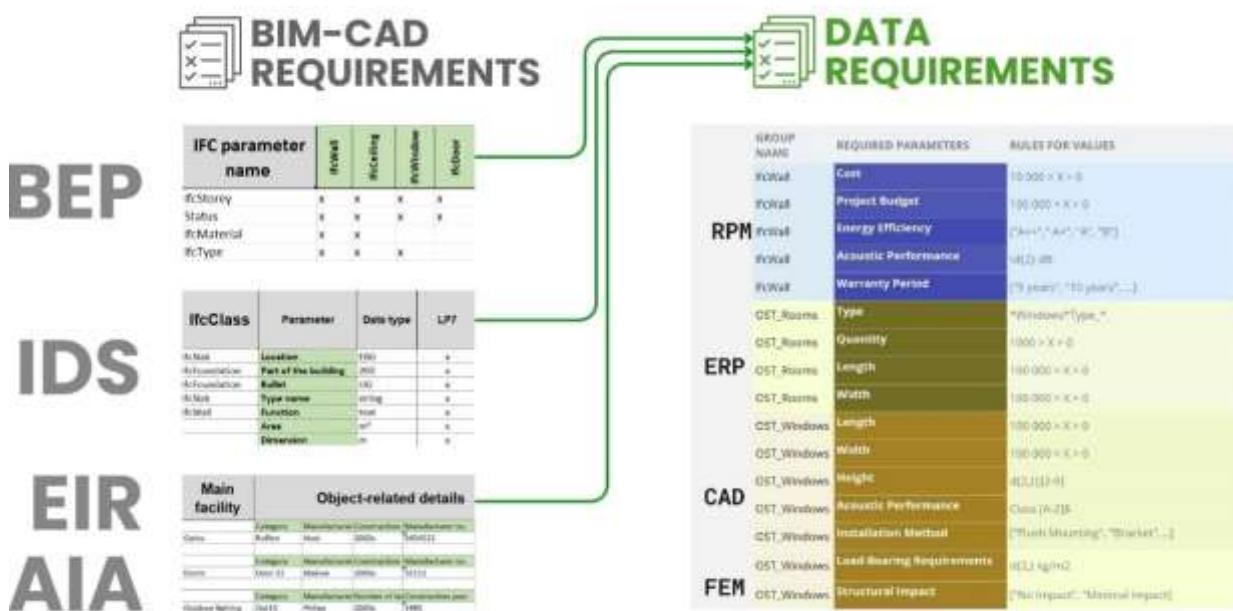
Az adatfeldolgozás egyszerűsítése helyett a CAD (BIM) adatokkal kapcsolatos új fogalmak a legtöbbször már az értelmezés és az alapvető meghatározások szakaszában további bonyolultságot és vitákat generálnak.

Az új koncepciók egyik legújabb példája az IDS formátum (2020-ban vezették be), amely lehetővé teszi az információs modell attribútum-összetételére vonatkozó követelmények leírását a nyílt BIM koncepcióban. Az IDS-követelmények az attribútumokra és azok határértékeire vonatkozó információkat strukturált táblázat formájában írják le (Excel vagy MySQL), amelyet aztán egy félig strukturált XML-formátum jelölésébe fordítanak le, amelyet XML-ről IDS speciális rövidítésre neveztek át.

A gyártók által támogatott és a BIM és a nyílt BIM által támogatott nézettel ellentében, miszerint az építőiparban az adatkezelés a speciális eszközök, például a CAD és a BIM használata miatt egyedülálló, az adatformátumok és adatkezelési gyakorlatok ebben az iparágban nem különböznek más iparágakétől.

A projektekre és a CAD (BIM) formátumokra vonatkozó követelmények száma egyszerűsíthető egyetlen, attribútum-oszlopokkal ellátott követelménytáblázat használatával, amelyet a "Követelmények strukturált formába történő lefordítása" című fejezetben részletezünk, anélkül, hogy az eredetileg strukturált követelményeket nem táblázatos formátumokra kellene lefordítani (az IDS eredetileg egy táblázat seqítségével kerül leírásra).

Az egyszerűsített megközelítés (6.3-2. ábra), amely tartalmazza a korábbi fejezetekben (4.4-9. ábra, 4.4-16. ábra, 7.3-10. ábra) részletesen tárgyalt entitások azonosítóit, tulajdonságait és határértékeit tartalmazó oszlopokat, kiküszöböli a követelmények IDS-XML formátumba történő átalakításának szükségeségét. Ez a módszer közvetlen, kevésbé nehézkes és átláthatóbb mechanizmust biztosít az adatminőség-ellenőrzéshez. Széles körben használt eszközökre támaszkodik, a reguláris kifejezésektől (RegEx) az adatkeretekig, a Pandasig és a szabványos ETL -payplanes - pontosan olyanokra, amelyeket a gazdaság más ágazataiban dolgozó szakemberek használnak az adatokkal való munkához.



6.3-2. ábra Más iparágakban az adatszolgáltatási követelmények az attribútumok és határértékeik strukturált leírására egyszerűsödnek.

Az építőiparban az adatok zárt jellege miatt idővel egyre több új megközelítés és technika jelenik meg a sokféle adat ellenőrzésére és kezelésére, bár az építési projektek adatai lényegében ugyanazok, mint más területeken. Míg más iparágak sikeresen beérík az adatfeldolgozás szabványosított megközelítéseivel, az építőiparban folyamatosan új és egyedi adatformátumok, követelmények és érvényesítési koncepciók kerülnek kidolgozásra.

Az építőiparban az adatok gyűjtésére, előkészítésére és elemzésére használt módszerek és eszközök nem különbözhetnek alapvetően azoktól, amelyeket a gazdaság más ágazataiban dolgozó szakemberek használnak.

Az iparágban kialakult egy sajátos terminológiai ökoszisztema, amely kritikus mérlegelést és újraértékelést igényel:

- A STEP formátum az új IFC név alatt helyezkedik el, kiegészítve az építési kategorizálással, anélkül, hogy figyelembe venné magának a STEP formátumnak a korlátait.
- Az IFC parametrikus formátumot használják az adatkommunikációs folyamatokban annak ellenére, hogy nincs egységes geometriai mag, amely a vizualizációhoz és a számításhoz szükséges.
- A CAD -rendszer adatbázisaihoz való hozzáférést a "BIM" kifejezés alatt népszerűsítik, anélkül, hogy az adatbázisok sajátosságait és az azokhoz való hozzáférést tárgyalnák.
- A szállítók az IFC és USD formátumokon keresztül népszerűsítik az átváratáságot, gyakran anélkül, hogy a gyakorlatban megvalósítanák azokat, költséges visszafejlesztést alkalmazva, amellyel ők maguk is küzdenek.
- A LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD kifejezéseket általánosan használják ugyanazon egységeket leírására, a más iparágakban régóta használt modellezési és ellenőrzési eszközökre való hivatkozás nélkül.

Az építőipar azt mutatja, hogy a fentiek - bár furcsán hangzik - lehetségesek az építőiparban, különösen, ha a fő cél az adatfeldolgozás minden egyes szakaszának pénzzé tétele a speciális szolgáltatások és szoftverek értékesítésén keresztül. Üzleti szempontból nincs ezzel semmi baj. Az azonban, hogy a CAD-hez (BIM) kapcsolódó ilyen rövidítések és megközelítések valóban hozzáadták értéket teremtenek-e és egyszerűsítik-e a szakmai folyamatokat, továbbra is nyitott kérdés marad.

Az építőiparban egy ilyen rendszer azért működik, mert maga az iparág a spekulatív nyereség nagy részét a rendszerek és rövidítések útvesztőjében szerzi. Az átlátható folyamatokban és a nyílt adatokban érdekelt vállalatok ritkák. Ez az összetett helyzet valószínűleg a végtelenségig fennmarad - mindaddig, amíg a vevők, az ügyfelek, a befektetők, a bankok és a magántőke nem kezdik el követelni az információkezelés világosabb és tájékozottabb megközelítését.

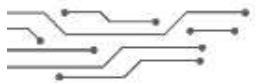
Az iparágban túlzottan sok rövidítés halmozódott fel, de ezek mind ugyanazokat a folyamatokat és adatigényeket írják le, különböző mértékben. Valódi hasznosságuk a munkafolyamatok egyszerűsítésében továbbra is kérdéses.

Miközben a fogalmak és a marketinges rövidítések jönnek és mennek, maguk az adatkövetelmények érvényesítési folyamatai örökre az üzleti folyamatok szerves részei maradnak. Ahelyett, hogy a egyre több és több speciális formátumot és szabályozást hozna létre, az építőiparnak olyan eszközökkel kellene támaszkodnia, amelyek más területeken, például a pénzügy, az ipar és az IT területén már beváltak.

A rengeteg kifejezés, rövidítés és formátum a mélyen kidolgozott digitális építési folyamatok illúzióját kelti. A marketingfogalmak és a bonyolult terminológia azonban gyakran egyszerű, de kényelmetlen igazságot takar: az adatok továbbra is nehezen hozzáférhetők, rosszul dokumentáltak és mereven meghatározott szoftvermegoldásokhoz kötöttek.

Ahhoz, hogy kilépjünk a rövidítések és a formátumok ördögi köréből, a CAD (BIM) rendszerekre nem

varázslatos információkezelési eszközként kell tekintenünk, hanem úgy, mint amik valójában - speciális adatbázisok. És ezen a prizmán keresztül lehet megérteni, hol végződik a marketing és hol kezdődik a valódi információval való munka.



FEJEZET 6.4.

TERVEZÉSI PARAMÉTEREZÉS ÉS AZ LLM HASZNÁLATA A CAD ÜZEMELTETÉSHEZ

A CAD-adatok egyediségének illúziója (BIM): az analitika és a nyílt formátumok felé vezető út.

A modern CAD (BIM) platformok jelentősen átalakították a tervezés és az építési információkezelés megközelítését. Míg korábban ezeket az eszközöket elsősorban rajzok és 3D modellek készítésére használták, ma már teljes értékű projektadattárként működnek. Az egységes igazságforrás koncepció szerint a parametrikus modell egyre inkább a projektinformációk fő és gyakran egyetlen forrásává válik, biztosítva annak integritását és relevanciáját a projekt teljes életciklusa során.

A legfontosabb különbség a CAD - (BIM -) platformok és más építőipari adatkezelő rendszerek között az, hogy az információkhoz (az egyetlen igazságforráshoz) való hozzáféréshez speciális eszközökre és API-kra van szükség. Ezek az adatbázisok nem univerzálisak a hagyományos értelemben: a nyílt struktúra és a rugalmas integráció helyett zárt környezetet alkotnak, egy adott platformhoz és formátumhoz kötöttek.

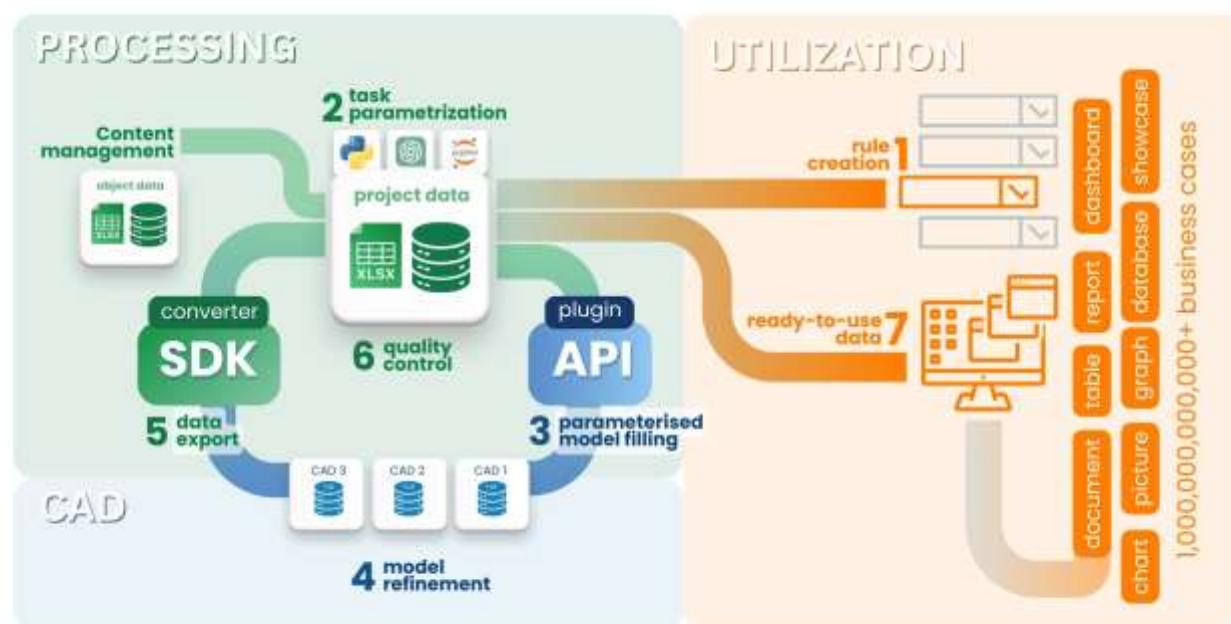
A CAD -adatokkal való munka összetettsége ellenére van egy fontosabb kérdés, amely túlmutat a technikai megvalósításon: mi is valójában a CAD-adatbázisok (BIM)? E kérdés megválaszolásához túl kell lépni a szoftverfejlesztők által előírt szokásos rövidítésekben és fogalmakon. Ehelyett érdemes a projektinformációkkal való munka lényegére összpontosítani: az adatokra és azok feldolgozására.

Az építőiparban az üzleti folyamat nem a CAD - vagy a BIM - eszközökben végzett munkával kezdődik, hanem a projektkövetelmények kialakításával és az adatmodellezéssel. Először a feladat paramétereit határozzák meg: az entitások listáját, azok kezdeti jellemzőit és határértékeit, amelyeket figyelembe kell venni egy adott feladat megoldása során. Csak ezután jönnek létre a modellek és elemek a CAD (BIM) rendszerekben a meghatározott paraméterek alapján

A CAD - (BIM-) adatbázisokban az információk létrehozását megelőző folyamat teljesen megegyezik az adatmodellezési folyamattal, amelyet a könyv negyedik részében és az "Adatmodellezés: fogalmi, logikai és fizikai modell" című fejezetben részletesen tárgyaltunk (4.3-1. ábra).

Ahogy az adatmodellezés során a követelményeket állítjuk össze az adatbázisban később feldolgozni kívánt adatokra vonatkozóan, a CAD-adatbázisok esetében a tervezési követelményeket a tervező több táblázatoszlop vagy kulcs-érték párok listája formájában állítják össze (6.4-1. ábra, 1-2. lépés). És csak ezen kezdeti paraméterek alapján az API automatikus vagy manuális használatával a tervező a CAD- (BIM) adatbázisokban objektumokat hoz létre (vagy inkább finomít) (3-4. lépések), majd ezeket ismét ellenőri a kezdeti követelményeknek való megfelelés szempontjából (5-6. lépések). Ez a folyamat - meghatározás→ létrehozás→ érvényesítés→ kiigazítás (2-6. lépés) - iteratív módon ismétlődik, amíg az adatminőség - az adatmodellezéshez hasonlóan - el nem éri a célrendszer - dokumentumok, táblázatok

vagy műszerfalak - számára kívánt szintet (7. lépés).

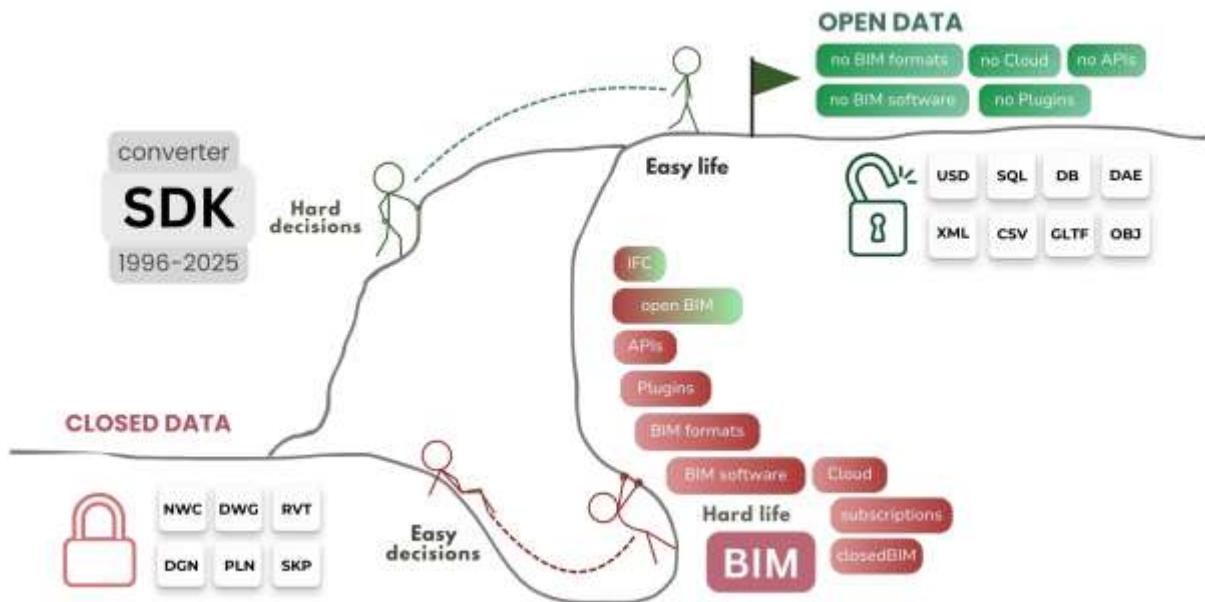


6.4-1 ábra Az adatbázisok információval való telítettségének ciklusa az üzleti folyamatok számára az építési projektek megvalósításában.

Ha a CAD (BIM) a tervezési környezeten kívül meghatározott követelmények alapján generált kulcs-érték párok formájában történő paraméterátvitel mechanizmusának tekintjük (6.4-1. ábra, 1-2. lépés), akkor a vita középpontja a konkrét szoftvermegoldásokról és azok korlátairól az alapvetőbb szempontokra - az adatszerkezetre, az adatmodellekre és az adatkövetelményekre - tevődik át. Lényegében az adatbázis paramétertelítettségről és a klasszikus adatmodellezési folyamatról van szó (2-3. és 5-6. lépés). Az egyetlen különbség az, hogy a CAD-adatbázisok zárt jellege és a használt formátumok sajátosságai miatt ezt a folyamatot speciális BIM-eszközök használata kíséri. Felmerül a kérdés: miben rejlik a BIM egyedisége, ha más iparágakban nincsenek hasonló megközelítések?

Az elmúlt 20 évben a BIM több mint egy egyszerű adatforrásként jelent meg. A CAD -BIM csomagot gyakran úgy forgalmazzák, mint egy olyan parametrikus eszközt, amely eredendően integrált adatbázissal rendelkezik [64], és amely képes automatizálni az építési projektek tervezési, modellezési és életciklus-menedzsment folyamatait. A valóságban azonban a BIM inkább eszközévé vált, hogy a felhasználókat a gyártók platformján tartsa, mintsem az adat- és folyamatmenedzsment kényelmes módszerévé.

Ennek eredményeképpen a CAD- (BIM-) adatokat elszigetelik a platformjaikon belül, és a projektinformációkat a saját API-k és geometriamagok mögé rejtik. Ez megfosztotta a felhasználókat attól a lehetőségtől, hogy a szállítói ökoszisztemákat megkerülve önállóan hozzáférjenek a adatbázisokhoz, és adatokat vonjanak ki, elemezzék, automatizáljanak és továbbítsanak más rendszerekbe.



Ábra 6.4-2 Az építőiparban a modern formátumok kifinomult geometriai magokat igényelnek, évente frissített API-t és speciális licenceket a CAD -(BIM-) szoftverekhez.

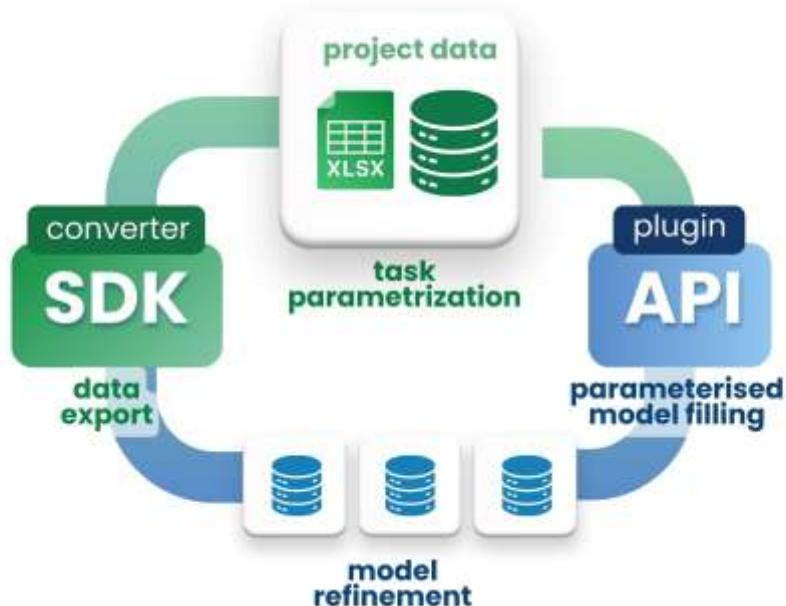
A modern CAD-eszközökkel dolgozó vállalatoknak ugyanazt a megközelítést kell alkalmazniuk az adatokkal való munkához, amelyet a gyakorlatban kivétel nélkül maguk a CAD-gyártók is alkalmaznak: az SDK - reverse engineering tools - segítségével történő adattranszformációt, amely ellen a CAD-gyártók 1995 óta küzdenek [75]. A CAD-adatbázishoz való teljes hozzáférés birtokában és a reverse engineering eszközök használatával [127] attribútumokkal ellátott entitások sík halmazát kaphatjuk meg, és exportálhatjuk őket bármilyen kényelmes nyílt formátumba (6.4-2. ábra), beleértve a tervezési elemek geometriáját és paramétereit is. Ez a megközelítés alapvetően megváltoztatja az információval való munka paradigmáját - a fájlközponttól az adatközpontú architektúráig:

- Az olyan adatformátumok, mint az RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET és mások azonos információkat tartalmaznak ugyanazon projekt elemeiről. Ez azt jelenti, hogy egy adott formátum és annak sémájának ismerete nem lehet akadálya az adatokkal való munkának.
- Bármilyen formátumból származó adatok egyetlen nyitott, strukturált és szemcsés struktúrában kombinálhatók (9.1-10. ábra), amely tartalmazza a MESH háromszöggeometriát és az összes objektum entitás tulajdonságait, a geometriai magok megkötéseivel nélkül.
- Az adatalemzés egyetemesre törekszik: a nyílt adatok használatával a projektadatokkal dolgozhat, függetlenül a használt formátumtól.
- Minimalizálás, valamint a API-któl és a szállítói bővítményektől való függőség: az adatokkal való munka többé nem függ az API-ismeretekről.

Amikor a és a CAD -adatigényeket könnyen elemezhető strukturált ábrázolási formátumokba alakítják át - a fejlesztők többé nem függnek a specifikus adatsémáktól és a zárt ökoszisztemáktól.

Tervezés paramétereiken keresztül: a CAD és a BIM jövője.

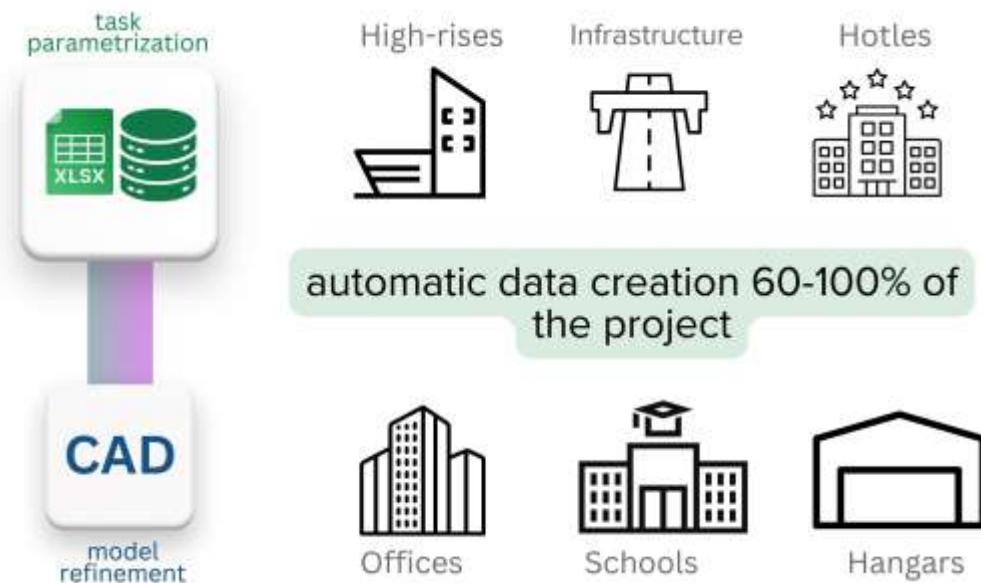
A világon még egyetlen építési projekt sem indult CAD programmal. Mielőtt egy rajz vagy modell a CAD-ben formát ölt, átmegy a koncepcióalkotási szakaszban (6.4-1. ábra, 1-2. szakasz), ahol a hangsúly azokon a paramétereken van, amelyek meghatározzák a leendő objektum alapötletét és logikáját. Ez a szakasz megfelel az adatmodellezés koncepcionális szintjének (4.3-6. ábra). A paraméterek létezhetnek kizártólag a tervező elméjében, de ideális esetben strukturált listák, táblázatok vagy adatbázisokban tárolt adatok formájában vannak elrendezve (6.4-3. ábra), ami lehetővé teszi az átláthatóságot, a reprodukálhatóságot és a tervezési folyamat további automatizálását.



6.4-3 ábra A tervezési folyamat egy iteratív folyamat, amelynek során a CAD-adatbázist kívülről származó információkkal töltjük fel az értéklánc követelményeinek felhasználásával.

Magának a CAD-modellezésnek a megkezdése előtt (az adatmodellezés logikai és fizikai fázisa (4.3-7. ábra)) fontos meghatározni a projekt alapjáról szolgáló határvonalparamétereket. Ezeket az attribútumokat - a többi követelményhez hasonlóan - az adatfelhasználási lánc legvégétől (pl. rendszerek) gyűjtik össze, és ezeken keresztül már meghatározásra kerülnek a projektben szereplő jövőbeli objektumok korlátai, céljai és fő jellemzői.

Maga a modellezés 60-100%-ban teljesen automatizálható a parametrikus modellező eszközök segítségével (6.4-3. ábra), ha a követelmények jól meghatározottak. Amint a projekt paraméterek formájában leírásra kerül, annak kialakítása technikailag megvalósíthatóvá válik, például a modern CAD - környezetekbe ágyazott vizuális programozási nyelvek, mint a Grasshopper Dynamo, vagy a Blender, UE, Omniverse ingyenes megoldások segítségével.



6.4-4 A tipizált projektek többsége ma már teljesen automatikusan készül a parametrikus programozóeszközöknek köszönhetően.

A nagy ipari és tipizált projekteket már ma sem a tervezési osztály keze által, hanem parametrikus eszközökkel és vizuális programozással hozzák létre. Ez lehetővé teszi, hogy egy adott tervező vagy menedzser szubjektív döntései helyett adatokon alapuló modellt építsenek.

A tartalom megelőzi a tervezést. A tartalom nélküli tervezés nem tervezés, hanem dekoráció [128].

- Jeffrey Zeldman, webdesigner és vállalkozó

A folyamat nem a rajzolással vagy a 3D -modellezéssel kezdődik, hanem a követelmények kialakításával. A követelmények határozzák meg, hogy a projektben milyen elemeket fognak használni, milyen adatokat kell átadni más részlegeknek és rendszereknek. Csak a strukturált követelmények megléte teszi lehetővé a modellek rendszeres automatikus ellenőrzését (például akár 10 percenként is, anélkül, hogy a tervezőt elvonná a munkájától).

Talán a jövőben a CAD- (BIM-) rendszer csak egy interfész lesz az adatbázis feltöltéséhez, és nem számít majd, hogy a modellezés melyik CAD-eszközben (fizikai szinten).

Hasonlóképpen, a gépészetben is gyakran használják a 3D modellezést, de ez nem szükséges vagy kötelező eleme a projektnek. A legtöbb esetben elegendő a klasszikus 2D-s dokumentáció, és annak alapján készül el a szükséges információs modell. Ez a modell az ipari szabványok szerint strukturált összetevőkből áll össze, és tartalmazza a tervezés és a gyártásszervezés megértéséhez szükséges összes információt. A gyári információs modellből ezután egy gyári információs modell készül, amelyhez konkrét, már a technológusok igényeihez igazodó termékeket és folyamatábrákat adnak hozzá. Az egész folyamat

felesleges bonyolultság nélkül szervezhető meg, anélkül, hogy a rendszert túlterhelnénk 3D grafikákkal, ahol azok nem nyújtanak valódi előnyöket.

Fontos megérteni, hogy maga a 3D modell és a CAD -rendszer nem játszhatja a főszerepet - ez csak egy eszköz a mennyiségi és geometriai elemzéshez. A geometria kivételével minden más paramétert, amely az entitást leírja, lehetőség szerint a CAD-környezeten kívül kell tárolni és feldolgozni (BIM).

A paramétereken keresztül történő tervezés nem csupán egy trend, hanem az építőipar elkerülhetetlen jövője. Az összetett 3D -modellek kézzel történő létrehozása helyett a tervezők adatokkal fognak dolgozni, validálni és automatizálni a folyamatokat, közelebb hozva az építőipart a programozás világához. Idővel a tervezési folyamatok a szoftverfejlesztés elveire épülnek majd:

- Követelmények létrehozása → Modell létrehozása→ Feltöltés a szerverre→ Ellenőrzés változások→ Pull request
- A Pull-kérelem automatikusan lefuttatja a modelellenőrzéseket a tervezési folyamat előtt vagy alatt létrehozott követelményekkel szemben.
- Az adatok minőségének ellenőrzése és jóváhagyása után a módosítások bevezetésre kerülnek a projektbe, a közös adatbázisba vagy automatikusan átkerülnek más rendszerekbe.

A gépészetiben az ilyen tervmódosítások már most is a változásbejelentés elkészítésével kezdődnek. Hasonló rendszer vár az építőiparra is: a tervezés iteratív folyamat lesz, ahol minden egyes lépést paraméteres követelményekkel támogatnak. Egy ilyen rendszer lehetővé teszi majd a tervezők számára, hogy automatizált ellenőrzéseket és automatizált pull-kérelmeket hozzanak létre a konkrét követelményekhez.

A jövő tervezője elsősorban adatoperátor, nem pedig kézi modellező. Az ő feladata, hogy a projektet paraméteres entitásokkal töltse fel, ahol a geometria csak az egyik attribútum.

Az átalakulásban fontos szerepet fog játszani az adatmodellezés, az osztályozás és a szabványosítás fontosságának megértése, amelyeket a könyv előző fejezeteiben részletesen tárgyaltunk. A jövő tervezési előírásait kulcs-érték paraméterpárokként formalizálják majd XLSX vagy XML -sémák formájában.

Az építőipar jövője az adatok gyűjtéséről, elemzéséről, validálásáról és a folyamatok automatizálásáról szól az analitikai eszközök segítségével. A BIM (vagy CAD) nem a végcél, hanem csak egy fejlődési szakasz. Amikor a szakemberek felismerik, hogy a hagyományos CAD-eszközöket megkerülve közvetlenül adatokkal dolgozhatnak, maga a "BIM" kifejezés fokozatosan átadja helyét a strukturált és szemcsés építési projektadatok használatának fogalmának.

Az átalakulást felgyorsító egyik legfontosabb tényező a nagy nyelvi modellek (LLM) és az ezeken alapuló eszközök megjelenése volt. Ezek a technológiák megváltoztatják a tervezési adatok kezelésének módját, lehetővé téve az információkhoz való hozzáférést anélkül, hogy a API-k vagy a szállítói megoldások mélyreható ismeretére lenne szükség. Az LLM-ek segítségével a követelmények létrehozásának folyamata és a CAD adatokkal való interakció intuitívvá és elérhetővé válik.

Az LLM megjelenése a tervezési CAD adatfeldolgozási folyamatokban.

A CAD-adatbázisokhoz való hozzáférési eszközök és a nyílt és egyszerűsített CAD -formátumok fejlesztése mellett az LLM -eszközök (Large Language Models) megjelenése forradalmasítja a tervezési adatok feldolgozását. Míg a múltban az információkhoz való hozzáférés főként bonyolult interfészeken keresztül történt, és programozási készségeket és az API ismeretét igényelte, ma már természetes nyelvi eszközökkel is lehet interakcióba lépni az adatokkal.

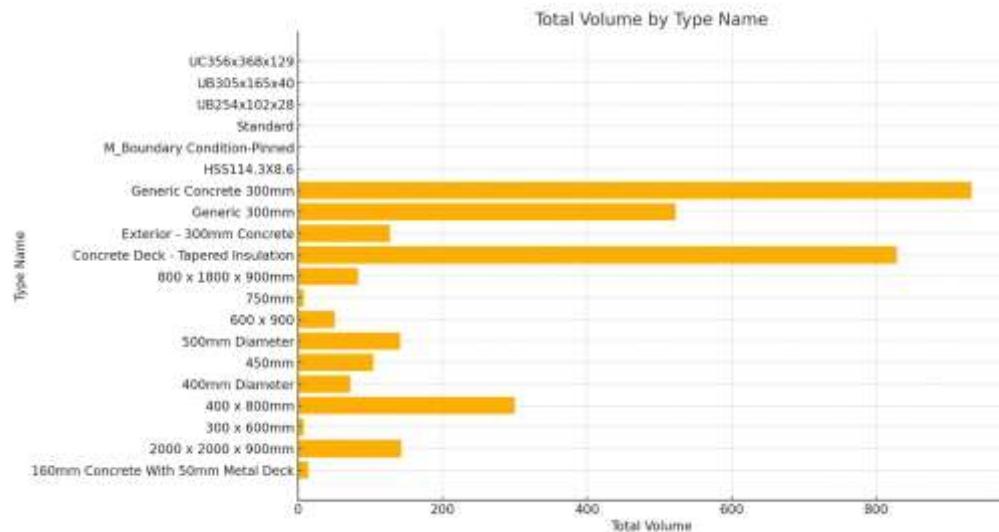
A műszaki háttérrel nem rendelkező mérnökök, menedzserek és tervezők a projektadatokból a szükséges információkat hétköznapi nyelven megfogalmazott lekérdezésekkel nyerhetik ki. Feltéve, hogy az adatok strukturáltak és hozzáférhetőek (4.1-13. ábra), elég egy kérdést feltenni az LLM csevegőszobában, például: "*Mutassa meg egy táblázatban típus szerinti csoportosítással a 10 köbméternél nagyobb térfogatú falakat*" - és a modell automatikusan átalakítja ezt a lekérdezést SQL vagy Pandas kóddá, létrehozva egy összefoglaló táblázatot, grafikont vagy akár egy kész dokumentumot.

Az alábbiakban néhány valós példát mutatunk be arra, hogy az LLM -modellek hogyan működnek együtt a különböző CAD- (BIM-) formátumokban megjelenített tervezési adatokkal.

- ❷ Példa egy LLM csevegésre egy RVT formátumú CAD-projektre irányuló lekérdezésre az RVT formátumba történő átalakítás után (4.1-13. ábra) egy táblázatos adatkeretbe (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vagy bármely más):

A RVT fájlból kapott adatkeret adatainak csoportosítása "Típusnév" szerint a "Volume" paraméter összegzésekor, és a csoportban lévő elemek számának megjelenítése. És kérjük, mutassa mindezt vízszintes hisztogramként, nulla értékek nélkül.

- ☒ LLM-válasz vízszintes oszlopdiagramként (PNG formátumban):



6.4-5 ábra 17 egérkattintás vagy 40 sornyí kód helyett a beépülő modulok használatával az LLM -ben azonnal lekérdezzük a QTO táblázatot egy szöveges lekérdezéssel.

- ☒ Ahhoz, hogy a "Falak" kategóriából létrehozzunk egy QTO táblázatot a falak típusairól, a teljes területtel és a mennyiséggel, fogalmazzunk meg egy szöveges lekérdezést az LLM -chat számára:

Vegye ki a projekt adatkeretből azokat az elemeket, amelyek "OST_Walls" paramétere "Category", csoportosítsa őket "Type Name" (típusnév) szerint, összegezze a "Area" (terület) oszlop értékét, adjon hozzá a mennyiséget, és a null értékek eltávolításával jelenítse meg őket a táblázatban.

- ☒ LLM válasz egy kész QTO táblázat formájában:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

6.4-6. ábra A QTO táblázat természetes nyelven történő létrehozása ugyanolyan minőségű eredményt ad, mint a CAD - (BIM-) eszközök használata esetén.

- ☒ Kérdezzük le a projektet IFC formátumban, miután táblázatos adatkeretbe konvertáltuk, és adjunk meg egy hasonló szöveges lekérdezést bármelyik LLM chatben:

Vegyük csak azokat az elemeket a projektből, amelyeknek a "Szülő" paraméterében az 1. és a 2. szint értékei szerepelnek, és vegyük azokat az elemeket, amelyeknek a "Kategória" paraméterében az IfcSlab értékei szerepelnek, majd csoportosítsuk ezeket az elemeket az "Objektumtípus" paraméter szerint, összegezzük az értékeket a "PSet_RVT _Dimensions Area" paraméterben, és jelenítsük meg őket kördiagramként.

☒ LLM válasz mint kész Tortadiagram az IFC adatokból származó elemcsoportokról :

For ■ Floor:127mm Slab on Grade, ■ Floor:150mm Exterior Slab on Grade,
 ■ Floor:Finish Floor - Ceramic Tile, ■ Floor:Finish Floor - Wood, and
 ■ Floor:Residential - Wood Joist with Subflooring



6.4-7 ábra Az IFC strukturált formátumú adatainak lekérdezésének eredménye bármilyen típusú grafikon lehet, amely alkalmas az adatok megértéséhez.

Az így kapott kész megoldások (6.4-5. ábra - 6.4-7. ábra) mögött egy tucatnyi sor Python kód áll, amely a Pandas könyvtárat használja. Az így kapott kód az LLM chatszobából kimásolható, és bármely helyi vagy online IDE-ben felhasználható, hogy az LLM chatszobán kívül is azonos eredményeket kapjunk.

Ugyanabban az LLM chatben nemcsak a 3D CAD (BIM) formátumból nyert projektekkel dolgozhatunk, hanem DWG formátumú sík rajzokkal is, amelyekre az LLM chatben lekérdezhetünk, hogy például az elemcsoportok adatait vonalak vagy 3D geometriák formájában jelenítsük meg, miután strukturált formába konvertáltuk őket.

DWG -fájlok automatizált elemzése az LLM és Pandas segítségével.

A DWG -fájlokból származó adatok feldolgozása - az információk strukturálatlansága miatt - mindenkor összetett feladat volt, amely speciális szoftvereket és gyakran kézi elemzést igényelt. A mesterséges intelligencia és az LLM eszközök fejlődésével azonban lehetővé vált számos lépés automatizálása, ez a ma még többnyire manuális folyamat. Nézzünk egy valódi csővezetéket az LLM-hez (ebben a példában a ChatGPT) intézett, a DWG-rajzokkal való munkavégzést lehetővé tevő kérésekről:

- DWG adatok szűrése réteg, azonosító és koordináták szerint.
- Az elemek geometriájának vizualizálása
- Automatikusan megjegyzésekkel látja el a rajzokat a paraméterek alapján
- A fal polivonalak kiterjesztése a vízszintes síkba
- Interaktív 3D -visualizációk létrehozása síkbeli adatokból

- Az építési adatok strukturálása és elemzése összetett CAD -eszközök nélkül.

Esetünkben a Pipeline építésének folyamata az LLM segítségével történő szekvenciális kódgenerálással kezdődik. Először egy, a feladatot leíró lekérdezést generálunk. A ChatGPT generálja a Python -kódot, amely végrehajtásra és elemzésre kerül, és az eredményt a csevegőszobán belül mutatja meg. Ha az eredmény nem a vártnak megfelelő, a kérést kijavítjuk, és a folyamatot megismételjük

A csővezeték az adatok feldolgozása és elemzése érdekében végzett automatizált lépések sorozata. Egy ilyen folyamat során minden egyes lépés bemenetként adatokat vesz fel, átalakításokat hajt végre, és az eredményt továbbítja a következő lépéseknek.

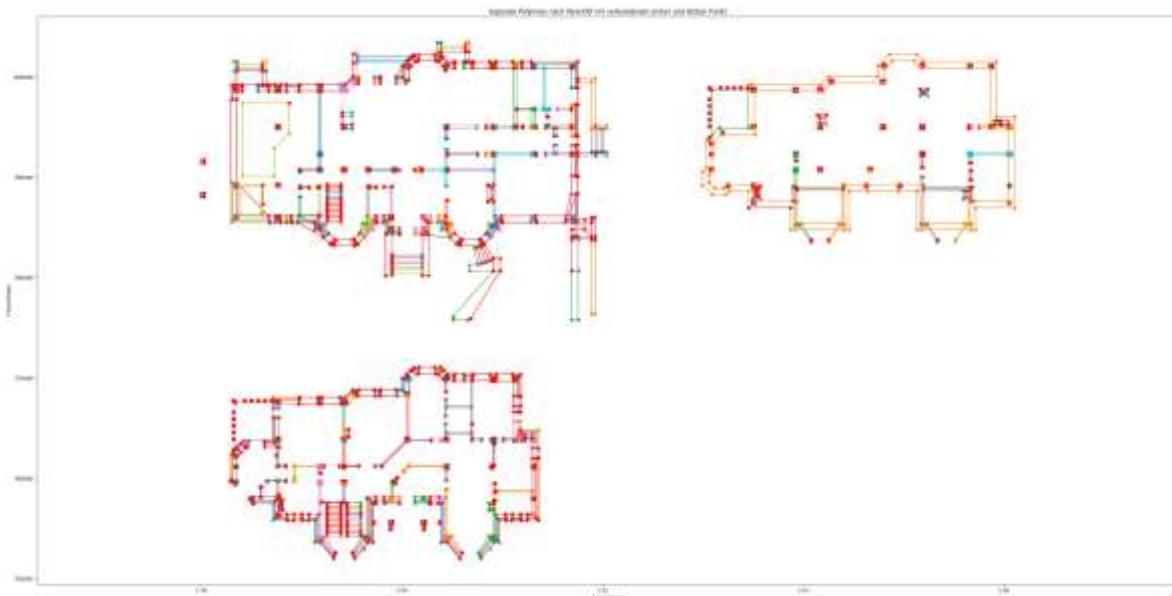
Miután megkaptuk a kívánt eredményt, a kódot kimásoljuk az LLM oldalról, és beillesztjük a kódba blokkok formájában bármelyik kényelmes IDE-ben, esetünkben a Kaggle platformon.com. Az így kapott kódrészleteket egyetlen Pipeline -ban egyesítjük, amely automatizálja a teljes folyamatot - az adatok betöltésétől azok végső elemzéséig. Ez a megközelítés lehetővé teszi az analitikai folyamatok gyors fejlesztését és skálázását mély programozási szakértelem nélkül. Az összes alábbi töredék teljes kódja, valamint a mintakérdések megtalálhatók a Kaggle.com platformon a "DWG Analyse with ChatGPT | DataDrivenConstruction" [129] kereséssel. [129].

Kezdjük a DWG adatokkal való munka folyamatát, a strukturált formába való átalakítás után (4.1-13. ábra), egy klasszikus lépéssel - az összes rajzadat csoportosításával és szűrésével, a feladatunkhoz szükséges falelemek, konkrétan a polivonalak (a "ParentID" paraméter lehetővé teszi a vonalak csoportosítását), amelyek a "Layer" paraméterben (adatkeret oszlop) a következő betűkombinációt (RegEx) tartalmazó string értékkel rendelkeznek - "wall".

- ▣ Ahhoz, hogy egy hasonló feladat kódját és az eredményt kép formájában megkapja, a következő lekérdezést kell írnia az LLM:

Először ellenőrizze, hogy a DWG-ből kapott adatkeret tartalmazza-e a meghatározott oszlopokat: ID', 'ParentID' és 'Point'. Ezután szűrje ki a "Layer" oszlobból azokat az azonosítókat, amelyek a "wall" karakterláncot tartalmazzák. Keresse meg a "ParentID" oszlobban azokat az elemeket, amelyek megfelelnek ezeknek az azonosítóknak. Határozzon meg egy függvényt a "Point" oszlop adatainak tisztítására és felosztására. Ez magában foglalja a zárójelek eltávolítását és az értékek "x", "y" és "z" koordinátákra való felosztását. Ábrázoljuk az adatokat a matplotlib segítségével. minden egyedi "ParentID" esetén rajzoljunk egy külön polivonalat a "Point" koordináták között. Győződjön meg róla, hogy az első és az utolsó pont lehetőleg össze van kötve. Állítsa be a megfelelő címeket és címeket, és gondoskodjon arról, hogy az x és y tengelyek egyformán méretezettek legyenek.

- ▣ A válasz LLM egy kész képet ad, amely mögött el van rejtve az azt generáló Python kód:



A 6.4-8 LLM kód a DWG fájlból kivette a "fal" réteg összes vonalát, kitisztította a koordinátáikat, és a Python könyvtárak egyikének segítségével megkonstruálta a polivonalakat.

- ❸ Most adjuk hozzá a vonalakhoz azt a területparamétert, amely minden egyes polivonal tulajdonságai között szerepel (az adatkeret egyik oszlopában):

Most kapunk csak egy "ParentID"-t minden egyes polivonalból - keressük meg ezt az ID-t az "ID" oszloban, vegyük a "Area" értéket, osszuk el 1,000,000-rel és adjuk hozzá ezt az értéket a grafikonhoz.

- Az LLM válasz egy új grafikont jelenít meg, ahol minden egyes polivonalhoz tartozik egy felirat a területével:

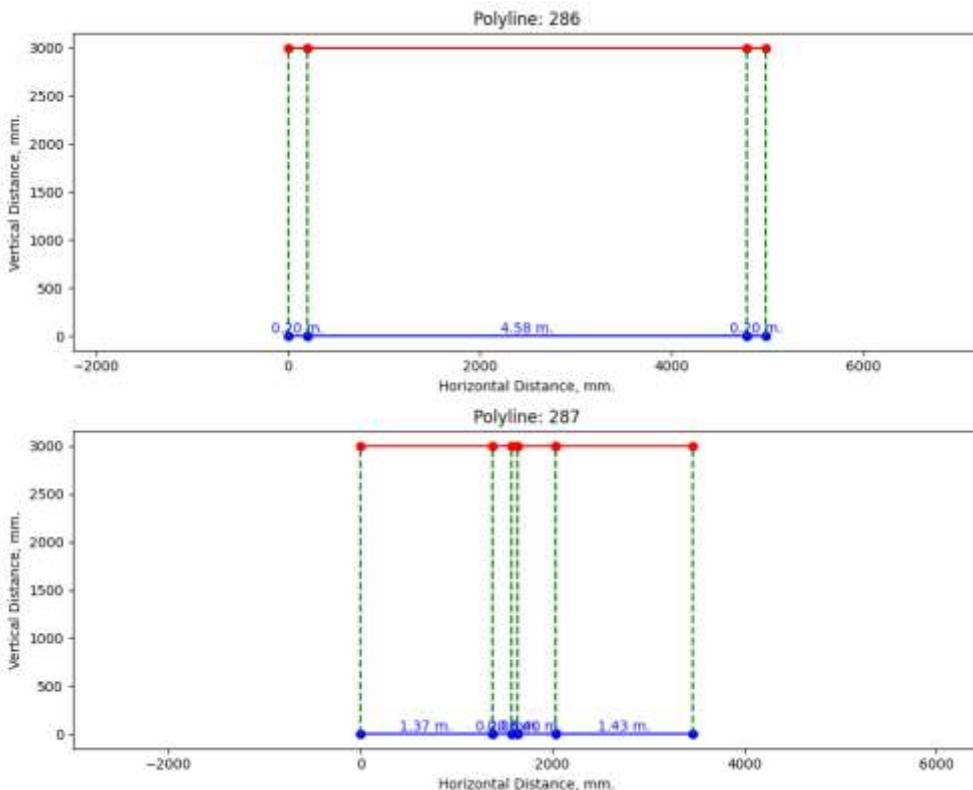


A 6.4-9. ábra Az LLM olyan kódot adott hozzá, amely az egyes polivonalak területértékeit veszi és hozzáadja a képhez a vonalak megjelenítésével.

- Ezután minden egyes polivonalat vízszintes vonallá alakítunk, hozzáadunk egy 3000 mm magasságú párhuzamos vonalat, és egyetlen síkba kapcsoljuk őket, hogy így mutassuk meg a falelemek felületeinek elrendezését:

A "Layer" oszlopból a "wall" értékkel rendelkező összes elemet ki kell venni. Vegyük ezeket az azonosítókat listaként az "ID" oszlopból, és keressük meg ezeket az azonosítókat a teljes adatkeretből a "ParentID" oszlopban. Az összes elem olyan vonal, amely egyetlen polivonalba van összevonva. minden vonalnak más x, y geometriája van, mint a "Point" oszloban szereplő első pontnak. minden egyes polivonalat sorban kell vennie, és a 0,0 ponttól vízszintesen meg kell rajzolnia a polivonal minden egyes szegmensének hosszát. a polivonal minden egyes szegmensének hosszát egy vonalba. Ezután pontosan ugyanezeket a vonalakat csak 3000-rel feljebb rajzolja, minden pontot egy síkba kapcsoljon össze.

- Az LLM válasz olyan kódot ad ki, amely lehetővé teszi a falrajzok síkban történő ábrázolását:

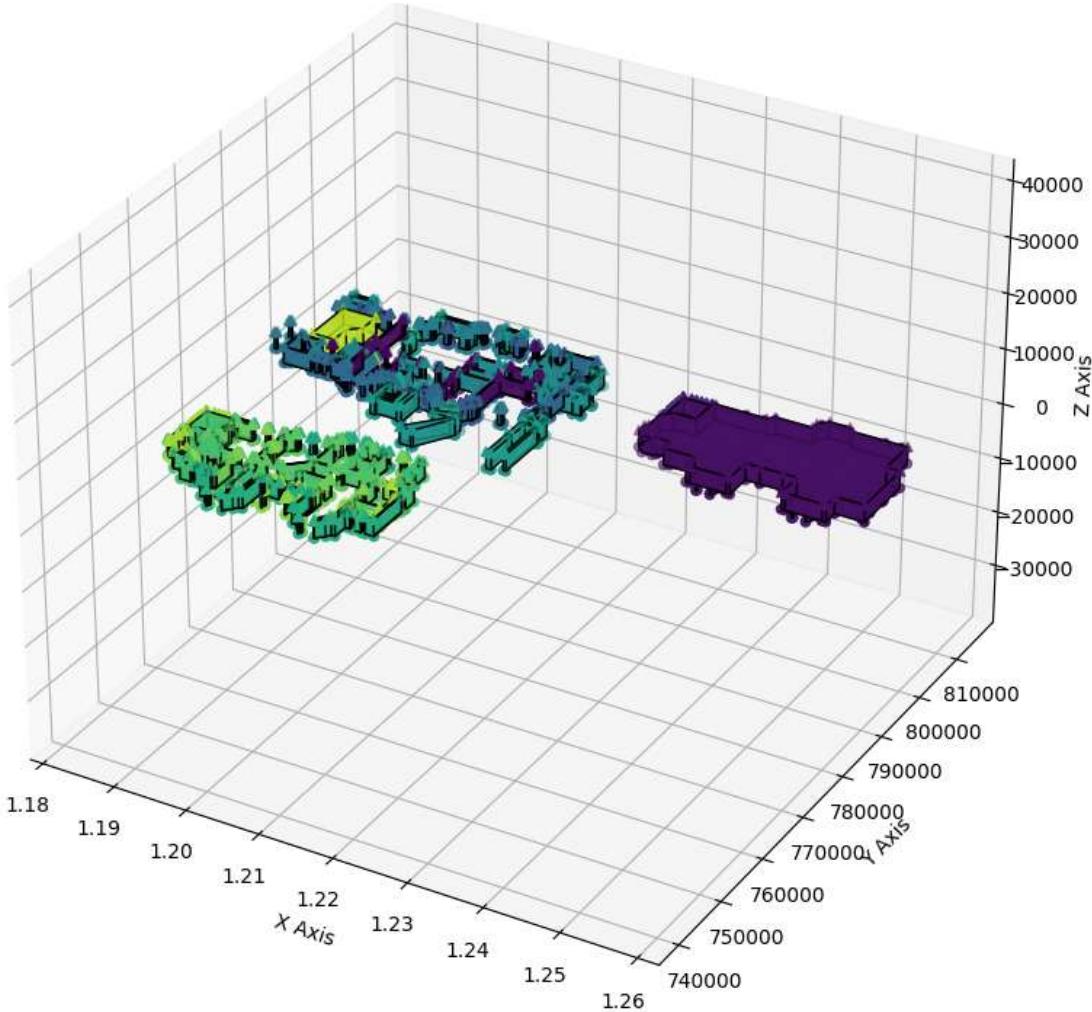


6.4-10. ábra minden egyes polivonalat felszólításokkal olyan elrendezéssé alakítunk, amely közvetlenül az LLM chatben vizualizálja a falsíkokat.

- Most pedig térjünk át a 2D-s vetítésről a 3D-s - a falakat sík vonalakból modellezük a polivonalak felső és alsó rétegének összekapcsolásával:

Jelenítse meg a falelemeket 3D-ben, összekötve a $z = 0$ és $z = 3000$ mm magasságú polivonalakat. Az épület falait ábrázoló zárt geometria létrehozása. Használja a Matplotlib 3D grafikus eszközt.

- Az LLM egy interaktív 3D -grafikát generál, amelyben minden egyes polivonal síkok halmazaként jelenik meg. A felhasználó szabadon mozoghat majd az elemek között a számítógép egere segítségével, 3D módban vizsgálva a modellt, a kódot a chatból az IDE-be másolva:



A 6.4-11 LLM segített kódot készíteni [129] a sík rajzvonalak 3D nézetére történő vizualizálásához, amely az IDE-n belül a 3D nézetben vizsgálható.

A logikus és reprodukálható Pipeline felépítéséhez - a DWG -fájl kezdeti konvertálásától és betöltésétől a végeredményig - ajánlott minden lépés után a generált LLM -kódblokkot az IDE-be másolni. Így nem csak a chatban ellenőrizheti az eredményt, hanem azonnal futtathatja a fejlesztőkörnyezetében. Ez lehetővé teszi, hogy a folyamatot szekvenciálisan építse fel, hibakereséssel és szükség szerinti kiigazítással.

Az összes töredék teljes Pipeline kódját (6.4-8. ábra 6.4-11. ábra 6.4-11. ábra) és a mintakérdéseket a Kaggle platformon.com a "DWG Analyse with ChatGPT | DataDrivenConstruction" kereséssel találja. [129]. A Kaggle-on nemcsak a kódot és a használt kéréseket lehet megtekinteni, hanem a teljes Pipeline-t az eredeti DWG adatkeretekkel a felhőben ingyenesen másolni és tesztelni is lehet, anélkül, hogy bármilyen további szoftvert vagy magát az IDE-t telepíteni kellene.

Az ebben a fejezetben bemutatott megközelítés lehetővé teszi a DWG -projekteken alapuló dokumentumok ellenőrzésének, feldolgozásának és generálásának teljes automatizálását. A kifejlesztett Pipeline alkalmas minden egyedi rajzok feldolgozására, minden több tíz, több száz és több ezer DWG-fájl kötegelt feldolgozására, a szükséges jelentések és vizualizációk automatikus generálásával minden egyes projekthez.

A folyamat szekvenciális és átlátható módon szervezhető: először a CAD -fájlból származó adatokat automatikusan XLSX formátumba konvertáljuk, majd betöltsük egy adatkeretbe, ezt követi a csoportosítás, az ellenőrzés és az eredménygenerálás - minden egyetlen Jupyter notebookban vagy Python -szkriptben valósítható meg, bárminely népszerű IDE-ben. Szükség esetén a folyamat könnyen kibővíthető a projektdokumentáció-kezelő rendszerekkel való integráció révén: a CAD-fájlok automatikusan lekérdezhetők megadott kritériumok szerint, az eredmények visszaküldhetők a tárolórendszerbe, a felhasználók pedig értesíthetők, ha az eredmények elkészültek - e-mailben vagy üzenetküldőn keresztül.

Az LLM csevegések és ügynökök használata a tervezési adatokkal való munkához csökkenti a speciális CAD -programuktól való függőséget, és lehetővé teszi az építészeti tervezés elemzését és vizualizálását anélkül, hogy a felülettel manuálisan kellene interakcióba lépni - egérkattintások és bonyolult menü navigáció megjegyzése nélkül.

Az építőiparban napról többet hallani az LLM, a granuláris strukturált adatokról, a DataFrames és az oszlopos adatbázisokról. A különböző adatbázisokból és CAD-formátumokból kialakított egységes kétdimenziós DataFrames ideális üzemanyag lesz a modern analitikai eszközök számára, amelyeket más iparágak szakemberei aktívan kezelnek.

Maga az automatizálási folyamat jelentősen leegyszerűsödik - ahelyett, hogy zárt, hiánypótló termékek API-ját tanulmányoznánk és bonyolult szkripteket írnánk a paraméterek elemzéséhez vagy átalakításához, mostantól elég lesz egy feladatot egyedi szöveges parancsok formájában megfogalmazni, amelyek a szükséges programozási nyelvhez szükséges Pipeline vagy Workflow-folyamatba illeszkednek, amely szinte bármilyen eszközön ingyenesen fut. Nem kell többé a CAD- (BIM-) eszközgyártók új termékeire, formátumaira, bővítményeire vagy frissítéseire várni. A mérnökök és az építők önállóan dolgozhatnak majd az adatokkal egyszerű, ingyenes és könnyen érthető eszközökkel, az LLM csevegések és ügynökök segítségével.

Következő lépések: a zárt formátumokról a nyílt adatokra való áttérés

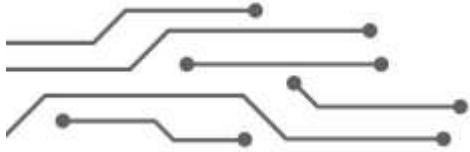
A jövő tervezési adataival való munka során nem valószínű, hogy bárki is szüksége lenne arra, hogy megértse a szabadalmaztatott eszközök geometriai magjait, vagy megtanuljon több száz, ugyanazt az információt tartalmazó inkompatibilis formátumot. Azonban anélkül, hogy megértenénk, miért fontos a nyílt strukturált adatokra való áttérés, nehéz érvelni az új ingyenes eszközök, nyílt adatok és olyan megközelítések használata mellett, amelyeket a szoftvergyártók valószínűleg nem támogatnak.

Ebben a fejezetben a CAD (BIM) adatok legfontosabb jellemzőit, korlátait és lehetőségeit ismertetünk, valamint azt, hogy a gyártók marketingigéretei ellenére a mérnökök és a tervezők minden nap nehézségekkel szembesülnek a tervezési információk kinyerése, átvitele és elemzése során. E rendszerek felépítésének megértése és az alternatív megközelítések megismerése - amelyek nyílt formátumokon és az LLM általi automatizáláson alapulnak - még egy-egy szakember, nem is beszélve a vállalatokról, életét is nagyban megkönnítheti. E rész összefoglalásaként érdemes kiemelni a főbb gyakorlati lépéseket, amelyek

segítségével az ebben a részben tárgyalt megközelítéseket alkalmazhatja a minden nap feladatai során:

- Bővítse a projektadatokkal való munkavégzéshez szükséges eszköztárát
 - Fedezze fel a rendelkezésre álló beépülő modulokat és segédprogramokat az adatok kinyeréséhez az Ön által használt CAD - (BIM-) rendszerekből.
 - Fedezze fel a rendelkezésre álló SDK-ket és API-kat, amelyek lehetővé teszik az adatok zárt formátumokból történő kinyerésének automatizálását anélkül, hogy manuálisan kellene speciális szoftvereket megnyitnia.
 - Alapvető késziségek elsajátítása a nyílt nem-parametrikus geometriai formátumokkal (OBJ, glTF, USD, DAE) és a megfelelő nyílt forráskódú könyvtárrakkal való munkavégzésben.
 - Próbáljon meg egy olyan rendszert kitalálni, amely a CAD (BIM) megoldásokon kívül a projekt metaadatait a geometriától elkülönítve tárolja, hogy egyszerűsítse az elemzést és a más rendszerekkel való integrációt.
 - Használja az LLM címet a formátumok közötti adatkonverziós problémák automatizálására.
- Hozzon létre saját folyamatokat a projektinformációk kezelésére
 - A feladatok és modellezési követelmények leírásának megkezdése paraméterek és értékeik segítségével, egyszerű és strukturált formátumban.
 - Személyes szkript- vagy kódblokkkönyvtár létrehozása a gyakran végzett műveletekhez
- A nyílt szabványok használatának előmozdítása munkája során
 - Kollégák és partnerek felkérése az adatok nyílt formátumban történő megosztására, amelyet nem korlátoz a szoftvergyártó ökoszisztemája.
 - A strukturált adatok használatának előnyeinek bemutatása konkrét példákon keresztül.
 - A zárt formátumokkal kapcsolatos problémákról és a lehetséges megoldásokról szóló megbeszélések kezdeményezése.

Még ha nem is tudja megváltoztatni a vállalatának a CAD - (BIM-) platformokkal kapcsolatos politikáját, a nyílt formátumú projektadatokkal való munka elveinek személyes megértése lehetővé teszi, hogy jelentősen növelte munkája hatékonyságát. A különböző formátumú adatok kinyerésére és átalakítására szolgáló saját eszközök és módszerek létrehozásával nemcsak a munkafolyamatokat optimalizálja, hanem rugalmasságot is nyer a szabványos szoftvermegoldások korlátainak megkerülésére.



VII RÉSZ

ADATVEZÉRELT DÖNTÉSHOZATAL, ANALITIKA, AUTOMATIZÁLÁS ÉS GÉPI TANULÁS

A hetedik rész az adatelemzésre és a folyamatok automatizálására összpontosít az építőiparban. Tárgyalja, hogyan válnak az adatok a döntéshozatal alapjává, és elmagyarázza az információk hatékony elemzéshez szükséges vizualizációjának elveit. Részletesen ismerteti a kulcsfontosságú teljesítménymutatókat (KPI), a beruházások megtérülésének (ROI) értékelési módszereit és a projektellenőrzéshez szükséges műszerfalak létrehozását. Különös figyelmet kapnak az ETL-folyamatok (Extract, Transform, Load) és azok automatizálása csővezetékek (Pipeline) segítségével, hogy a széttagolt adatokból strukturált információkat alakítsanak ki az elemzéshez. A munkafolyamatrendszerző eszközök, például az Apache Airflow, az Apache NiFi és az n8n, amelyek lehetővé teszik az automatizált adatpipeline-ok létrehozását mély programozási ismeretek nélkül. Jelentős szerepet játszanak a nagy nyelvi modellek (LLM) és azok használata az adatelemzés egyszerűsítésére és a rutinfeladatok automatizálására.

FEJEZET 7.1.

ADATELEMZÉS ÉS ADATVEZÉRELT DÖNTÉSHOZATAL

Az információk összegyűjtésének, strukturálásának, tisztításának és ellenőrzésének lépései után egy koherens és elemzhető adathalmaz jött létre. A könyv korábbi részei a heterogén források rendszerezésével és strukturálásával foglalkoztak - a PDF-dokumentumoktól és az ülések szöveges feljegyzéseitől a CAD-modellekig és geometriai adatokig. Részletesen ismertetjük az információk ellenőrzésének és a különböző rendszerek és osztályozók követelményeihez való igazításának, a duplikációk és következetlenségek kiküszöbölésének folyamatát.

Az ezeken az adatokon végzett összes számítás (a könyv harmadik, negyedik része) - az egyszerű transzformációtól az idő-, költség- és ESG-mutatók számításáig (ötödik rész) - aggregált elemzési feladatok. Ezek képezik az alapját egy projekt aktuális állapotának megértéséhez, paramétereinek értékeléséhez, majd a döntések meghozatalához. Ennek eredményeképpen az adatok a számítások eredményeként széttagolt rekordok halmazából olyan kezelhető erőforrássá válnak, amely képes megválaszolni a kulcsfontosságú üzleti kérdéseket.

Az előző fejezetek részletesen bemutatták az adatgyűjtési és minőségellenőrzési folyamatokat a tipikus üzleti esetekben és az építőiparra jellemző folyamatokban történő felhasználásra. Az analitika ebben a kontextusban sok tekintetben hasonlít más iparágak alkalmazásaihoz, de számos sajátossággal rendelkezik.

A következő fejezetek részletezik az adatelemzési folyamatot, beleértve az automatizálási lépéseket az információ kezdeti megszerzésétől és átalakításától a célrendszerekbe és dokumentumokba történő későbbi átvitelig. Először egy elméleti rész kerül bemutatásra, amely az adatvezérelt döntéshozatal kiválasztott szempontjaira összpontosít. Ezután a következő fejezetekben az ETL -Pipeline automatizálásához és felépítéséhez kapcsolódó gyakorlati rész következik.

Az adatok mint a döntéshozatalban használt erőforrás

Az adatvezérelt döntéshozatal gyakran iteratív folyamat, amely a különböző információforrásokból származó információk szisztematikus gyűjtésével kezdődik. A természetes körforgáshoz hasonlóan az egyes adatelemek és a teljes információs rendszerek fokozatosan a talajba hullanak - felhalmozódnak a vállalatok információs tárházaiban (1.3-2. ábra). Idővel ezek az adatok, mint a lehullott levelek és ágak, értékes anyaggá alakulnak át. Az adatmérnökök és elemzők micéliuma rendszerezi és előkészíti az információkat a jövőbeli felhasználásra, és a lehullott adatokat és rendszereket értékes komposzttá alakítja, hogy új hajtásokat és új rendszereket növesszen (1.2-5. ábra).

Az analitika széles körű használatának tendenciái különböző iparágakban egy olyan új korszak kezdetét jelzik, amelyben az adatokkal való munka a szakmai tevékenység alapjává válik (7.1-1. ábra). Fontos, hogy az építőipari szakemberek alkalmazkodjanak ezekhez a változásokhoz, és készen álljanak az új korszakba - az adatok és az analitika korszakába - való átlépésre

Az adatok táblázatok közötti kézi mozgatása és a számítások manuális elvégzése fokozatosan a múlté, és átadja helyét az automatizálásnak, az adatáramlás elemzésének, az analitikának és a gépi tanulásnak. Ezek az eszközök a modern döntéstámogató rendszerek kulcsfontosságú elemeivé válnak.

A McKinsey "Újraindítás. McKinsey's Guide to Overcoming Competition in the Age of Digital Technology and Artificial Intelligence" [130] idézi a különböző régiók, iparágak és funkcionális területek 1330 felsővezetőjének 2022-ben végzett felmérését. [130], idéz egy 2022-ben különböző régiók, iparágak és funkcionális területek 1330 felsővezetője körében végzett felmérést. Eredményei szerint a vezetők 70%-a fejlett analitikát használ saját ötleteinek generálására, 50%-uk pedig mesterséges intelligenciát alkalmaz a döntéshozatali folyamatok javítására és automatizálására.



ábra 7.1-1 Az adatelemzés és az analitika a legfontosabb eszköz a vállalaton belüli döntéshozatal sebességének növelésére.

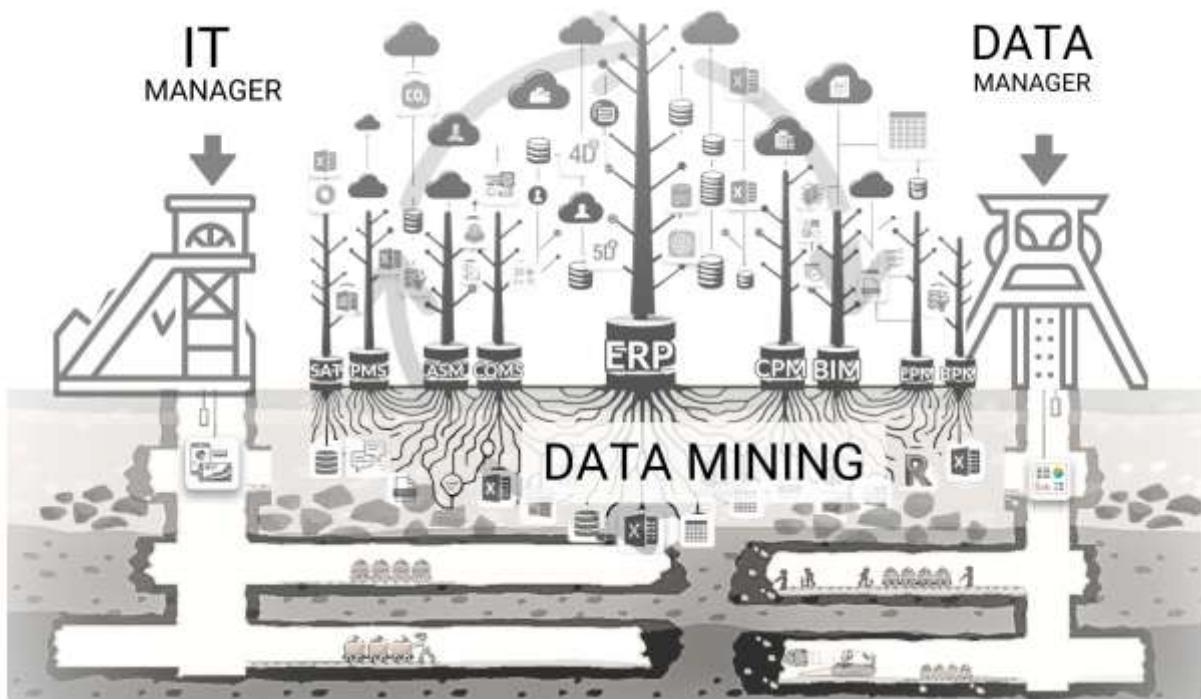
Az adatelemzés, mint a micélium terjedése, behatol a múltbeli döntések humuszába, segít összekapcsolni az egyes rendszereket, és értékes felismerésekhez vezeti a vezetőket. Ez a tudás, mint a korhadt adatrendszerek fáinak tápanyagai, új döntéseket táplál a vállalaton belül, ami hatékony változáshoz és minőségi információövekedéshez vezet, mint a gazdag és egészleges talajból kibújó új hajtások és csírák (1.2-5. ábra).

A számok fontos történetet mesélnek el. Számítanak rád, hogy világos és meggyőző hangot adj nekik [131].

- Stephen Few, adatvizualizációs szakértő

A közép- és kisvállalkozásoknál az információk kinyerése és további elemzésre való előkészítése ma rendkívül munkaigényes folyamat (7.1-2. ábra), amely a tizennyolcadik századi szénbányászathoz hasonlítható. A közelmúltig az adatbányászat és az adatelőkészítés inkább csak a kalendorok számára volt fenntartva, akik egy erősen specializált résben dolgoztak, és a strukturálattalan, lazán strukturált, vegyes és zárt forrásokból származó különböző típusú adatokkal való munkavégzéshez szűk és korlátozott eszköztárral rendelkeztek.

A döntéshozók és a vezetők gyakran nem rendelkeznek tapasztalattal a heterogén adatok és rendszerek kezelésében, mégis ezek alapján kell döntéseket hozniuk. Ennek eredményeként a modern építőiparban az elmúlt évtizedekben az adatvezérelt döntéshozatal kevésbé hasonlított egy automatizált folyamatra, inkább a bányászok többnapos kézi munkájára a korai szénbányákban.



Ábra: 7.1-2 Az adatbányászati folyamat során a szakértők az adatok előkészítésének összetett útját járják végig - a tisztítástól a későbbi elemzéshez szükséges strukturálásig.

Bár az építőiparban az adatok kinyerésének modern módszerei minden bizonnal fejlettebbek, mint a 12. századi bányászok primitív technikái, ez még mindig összetett és nagy kockázatú feladat, amely jelentős erőforrásokat és szakértelemet igényel, amit csak a nagyvállalatok engedhetnek meg maguknak. A adatoknak a múltbeli projektek felhalmozott örökségéből történő kinyerését és elemzését egészen a közelmúltig elsősorban olyan nagy, technológiaileg fejlett vállalatok végezték, amelyek évtizedek óta folyamatosan gyűjtik és tárolják az adatokat

Korábban az analitikában a vezető szerepet a technológiailag érett, évtizedek óta adatokat gyűjtő vállalatok játszották. Ma a helyzet változik: az adatokhoz és az adatfeldolgozó eszközök közékhöz való hozzáférés demokratizálódik - a korábban összetett megoldások ma már mindenki számára ingyenesen elérhetők.

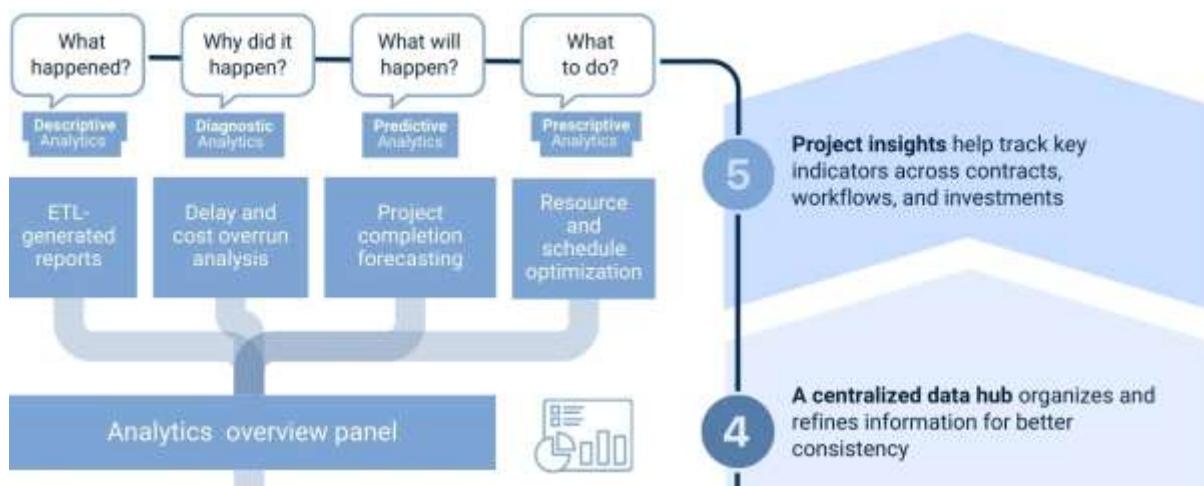
Az analitika alkalmazása lehetővé teszi a vállalatok számára, hogy pontosabb és megalapozottabb döntéseket hozzanak valós időben. A következő esettanulmány azt szemlélteti, hogy a múltbeli adatok hogyan segíthetnek a pénzügyileg megalapozott döntések meghozatalában:

- ⌚ **Projektmenedzser** - "Most a beton átlagára a városban 82€ /m³, a becslésben 95 €/m³ szerepel."
- ⌚ **Becslő** - "A korábbi projekteknél a túllépés körülbelül 15% volt, ezért visszaléptem."
- ⌚ **Adatmenedzser vagy ügyféloldali ellenőrző mérnök** - "Nézzük meg az utolsó három pályázat analitikáját".

A múltbeli projektek DataFrame elemzése után a következőket kapjuk:

- **Átlagos tényleges beszerzési ár:** 84,80 /m³€
- **Átlagos túlköltekezési arány:** +4,7%
- ⌚ **A becslésben ajánlott arány:** ~ 85 /m³€

Egy ilyen döntés már nem szubjektív érzéseken, hanem konkrét történelmi statisztikákon alapul, ami segít csökkenteni a kockázatokat és növeli a pályázati ajánlat érvényességét. A korábbi projektek adatainak elemzése egyfajta "szerves trágává" válik, amelyből új, pontosabb megoldások csíráznak ki.



7.1-3. ábra Az adatelemzés három kulcsfontosságú kérdésre ad választ: mi történt, miért történt, és mit kell tenni a továbbiakban.

A döntéshozók és a vezetők gyakran szembesülnek azzal, hogy heterogén adatokkal és rendszerekkel kell dolgozniuk megfelelő műszaki szakértelem nélkül. Ilyen helyzetekben a vizualizáció, amely az elemzési folyamat egyik első és legfontosabb lépése, kulcsfontosságú segítséget nyújt az adatok megértéséhez. Lehetővé teszi az információk vizuális és érthető formában történő bemutatását.

A adatok vizualizálása: a megértés és a döntéshozatal kulcsa

A mai építőiparban, ahol a projektadatokat a komplexitás és a többszintű struktúra jellemzi, a vizualizáció kulcsszerepet játszik. A adatok vizualizálása lehetővé teszi a projektmenedzserek és mérnökök számára, hogy a nagy, heterogén adatmennyiségekben elrejtett összetett mintákat és trendeket megjelenítsék.

A adatok vizualizálása megkönnyíti a projekt állapotának megértését: az erőforrások elosztása, a költségtendenciák vagy az anyagfelhasználás. A grafikonok és diagramok hozzáférhetővé és érthetővé teszik az összetett és száraz információkat, lehetővé téve a figyelmet igénylő kulcsfontosságú területek gyors azonosítását és a lehetséges problémák felismerését.

A adatok vizualizálása nemcsak az információk értelmezését könnyíti meg, hanem az elemzési folyamat és a megalapozott vezetői döntéshozatal egyik kulcsfontosságú lépése, amely segít megválaszolni a "mi történt?" és a "hogyan történt?" kérdéseket (2.2-5. ábra). (2.2-5. ábra).

A grafika a logikai problémák megoldásának vizuális eszköze [132].

- Jacques Bertin, "Grafika és grafikus információfeldolgozás

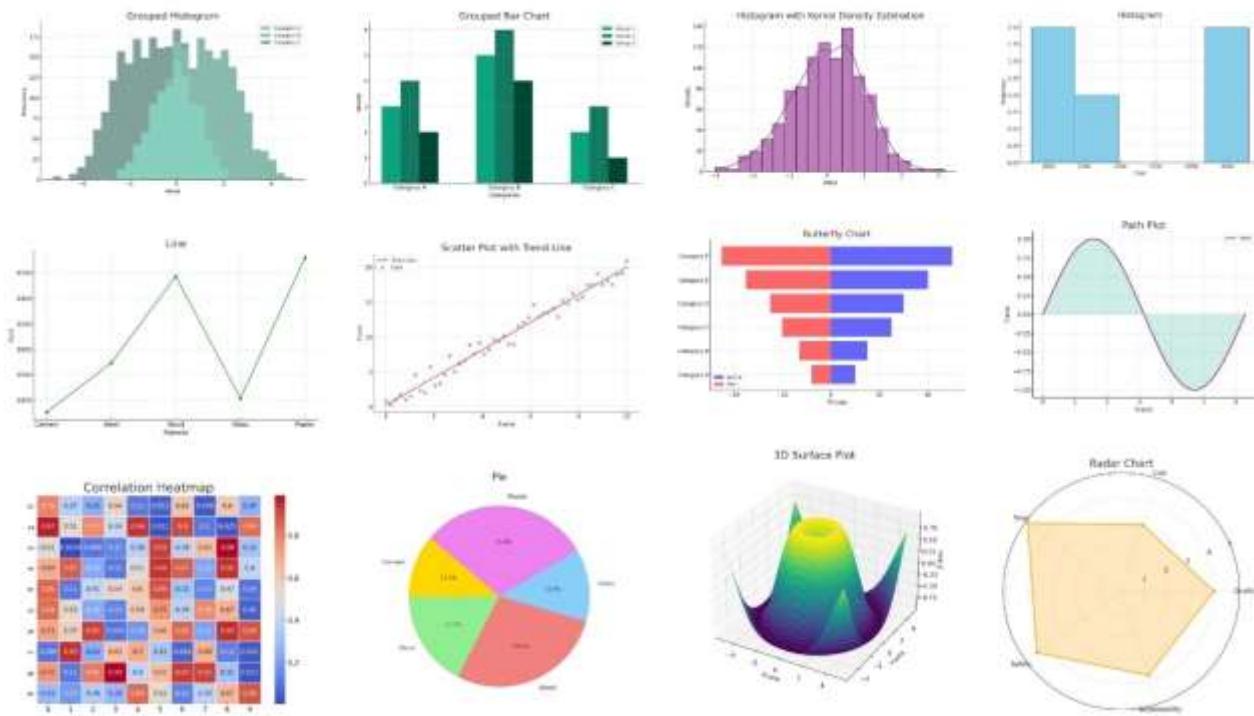
A kulcsfontosságú döntések meghozatala előtt a projektvezetők nagyobb valószínűséggel használják az adatok vizuális ábrázolását, mint a táblázatokból vagy szöveges üzenetekből származó száraz és nehezen értelmezhető számadatokat.

Az adatok vizualizáció nélkül olyanok, mint az építkezésen szerteszét heverő építőanyagok: a bennük rejő lehetőségek nem világosak. Csak akkor válik világossá az értékük, ha világosan vizualizáljuk őket, mint egy téglából és betonból épített házat. Amíg a ház meg nem épül, nem lehet megmondani, hogy az anyaghalomból kis kunyhó, luxusvilla vagy felhőkarcoló lesz-e.

A vállalatok különböző rendszerekből származó adatokkal (1.2-4. ábra - 2.1-10. ábra), pénzügyi tranzakciókkal és kiterjedt szöveges adatokkal rendelkeznek. Ezeknek az adatoknak az üzleti hasznosítása azonban gyakran kihívást jelent. Ilyen helyzetekben a vizualizáció fontos eszközzé válik az adatok jelentésének közvetítésében, segítve az információk bármely szakértő számára érthető formában, például műszerfalak, grafikonok és diagramok formájában történő bemutatását.

A PwC "What Students Need to Succeed in a Fast Changing Business World" (2015) című tanulmánya kiemeli [9], hogy a sikeres vállalatok az adatelemzésen túl aktívan használják az interaktív vizualizációs eszközöket, például grafikonokat, infografikákat és analitikus műszerfalakat a döntéshozatal támogatására. A jelentés szerint - az adatvizualizáció grafikonok, diagramok, műszerfalak és interaktív adatmodellek segítségével segít az ügyfeleknek megérteni az adatok által elmesélt történetet.

Az információk vizuális grafikai formákba, például diagramokba, grafikonokba és grafikonokba való átalakítása javítja az emberi agy megértését és az adatok értelmezését (7.1-4. ábra). Ez lehetővé teszi a projektmenedzserek és elemzők számára, hogy gyorsabban értékeljék az összetett forgatókönyveket, és megalapozott döntéseket hozzanak, amelyek az intuíció helyett vizuálisan felismerhető tendenciákon és mintákon alapulnak.



Ábra 7.1-4 A vizualizáció különböző típusai arra szolgálnak, hogy az emberi agy jobban megértse és értelmezze a számok száraz információt.

Az adatokból történő vizualizációk létrehozását és a különböző ingyenes vizualizációs könyvtárak használatát részletesebben az ETL -folyamatokról szóló következő fejezetben tárgyaljuk.

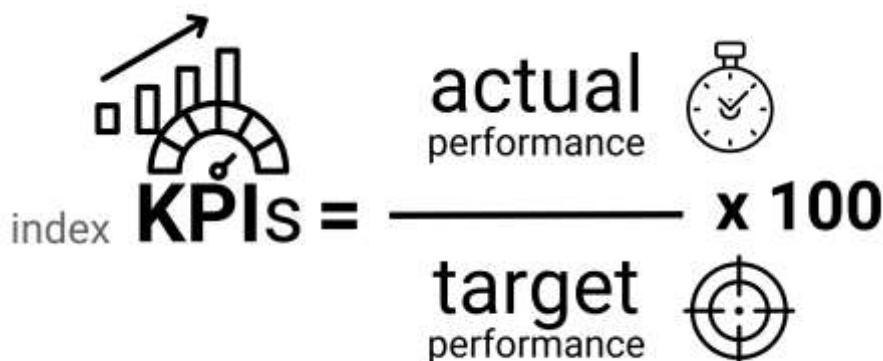
A vizualizáció az építőiparban az adatokkal való munka szerves részévé válik - nem csak az adatok "látását" segíti, hanem az adatok jelentésének megértését is az irányítási feladatok összefüggésében. Ahhoz azonban, hogy a vizualizáció valóban hasznos legyen, előzetesen meg kell határozni, hogy pontosan mit kell vizualizálni, és mely mérőszámok valóban fontosak a projektteljesítmény értékeléséhez. Itt jönnek a képhez az olyan teljesítménymérők, mint a KPI-k és a ROI. Ezek nélkül még a legszebb műszerfalnon is fennáll a veszélye annak, hogy csak "információs zaj" lesz.

KPI-k és ROI

A mai építőiparban a teljesítménymutatók (KPI és ROI) kezelése és azok jelentéseken és műszerfalakon keresztül történő megjelenítése kulcsfontosságú szerepet játszik a termelékenység és a projektmenedzsment hatékonyságának javításában.

Mint minden vállalkozásban, az építőiparban is világosan meg kell határozni azokat a mérőszámokat, amelyek alapján a sikert, a befektetés megtérülését és a teljesítményt mérik. A különböző folyamatokra vonatkozó adatok beszerzésekor egy adatközpontú szervezetnek először meg kell tanulnia meghatározni a kulcsfontosságú **teljesítménymutatókat (Key Performance Indicators)** - olyan mennyiségi mérőszámok, amelyek tükrözik a stratégiai és működési célok elérésének mértékét.

A KPI kiszámításához általában egy képletet használnak (7.1-5. ábra), amely tartalmazza a tényleges és a tervezett mutatókat. Például egy projekt, alkalmazott vagy folyamat egyedi KPI-jének kiszámításához ossza el a tényleges teljesítményt a tervezett teljesítménnyel, és szorozza meg az eredményt 100%-kal.



7.1-5. ábra KPI-k a projekt vagy folyamat sikerességének mérésére szolgálnak a fő célkitűzések elérésében.

Telephelyi szinten részletesebb KPI-ket lehet használni mérőszámok:

- **A kulcsfontosságú mérföldkövek** (alapozás, telepítés, befejezés) **időzítése** - lehetővé teszi a munkatervezek betartásának ellenőrzését.
- **Az anyagtúllépések százalékos aránya** - segít a beszerzés irányításában és a pazarlás minimalizálásában.
- **A nem tervezett gépleállások száma** - befolyásolja a termelékenységet és a költségeket.

A rossz mérőszámok kiválasztása téves "mit tegyünk?" döntésekhez vezethet (2.2-5. ábra). Ha például egy vállalat csak a négyzetméterenkénti költségekre összpontosít, de nem veszi figyelembe az átalakítások költségeit, az anyagokon való megtakarítás rosszabb minőséghoz és magasabb költségekhez vezethet a jövőbeli projekteknél.

A célok meghatározásakor fontos, hogy tisztában legyünk azzal, hogy mit mérünk. A homályos megfogalmazás téves következtetésekhez vezet és megnehezíti az ellenőrzést. Nézzünk példákat a sikeres és sikertelen KPI-kre az építőiparban.

Jó KPI-k:

- ☒ "Az év végére 10 százalékkal csökkentsük az újrafestési munkák arányát."
- ☒ "A következő negyedévre 15 százalékkal növeljük a homlozkati szerelés sebességét a minőség

romlása nélkül"

- ⌚ "Az év végére a munkarendek optimalizálásával 20%-kal csökkentjük a gépek állásidejét"

Ezek a mérőszámok egyértelműen mérhetőek, konkrét értékekkel és időkeretekkel rendelkeznek.

Rossz KPI-k:

- ⌚ "Gyorsabban fogunk építeni" (Mennyivel gyorsabban? Mit jelent a "gyorsabban"?).
- ⌚ "Javítani fogjuk a konkrét munka minőségét" (Pontosan hogyan mérik a minőséget?)
- ⌚ "Javítani fogjuk a vállalkozói interakciót a helyszínen" (Milyen kritériumok alapján lesz javulás?)

A jó KPI olyan, amely mérhető és objektíven értékelhető. Az építőiparban ez különösen fontos, mivel egyértelmű mutatók nélkül lehetetlen nyomon követni a teljesítményt és stabil eredményeket elérni.

A KPI mellett létezik egy további mérőszám a beruházások hatékonyságának értékelésére: a **ROI (Return on Investment)** - a befektetések megtérülésének mutatója, amely a nyereség és a befektetett pénzeszközök arányát tükrözi. A ROI segít annak értékelésében, hogy az új módszerek, technológiák vagy eszközök bevezetése indokolt-e: a digitális megoldásoktól és automatizálástól (pl. 7.3-2. ábra) az új építőanyagok alkalmazásáig. Ez a mutató segít megalapozott döntéseket hozni a további beruházásokról, azok üzleti jövedelmezőségre gyakorolt valós hatása alapján

Az építési projektmenedzsmentben a ROI (megtérülés) a kulcsfontosságú teljesítménymutatók (KPI) egyikeként használható, ha a vállalat célja egy projekt, technológia vagy folyamatfejlesztés megtérülésének mérése. Például, ha egy új építésmenedzsment-technika bevezetésére kerül sor, a ROI megmutathatja, hogy az mennyivel javította a nyereségességet.

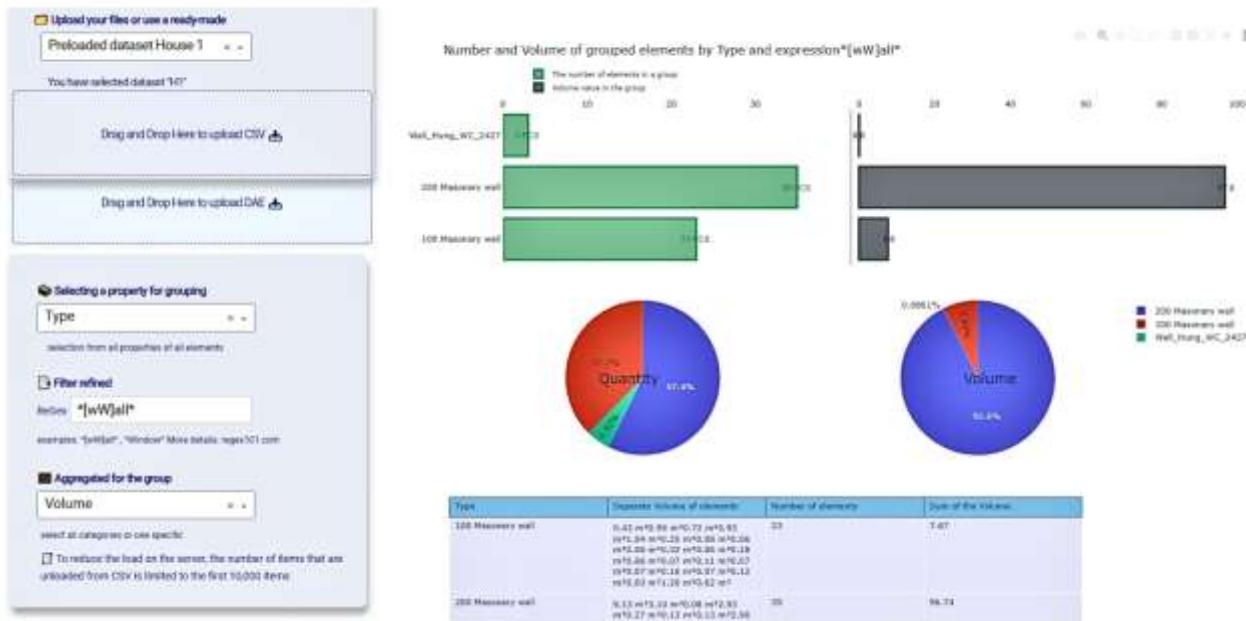
A KPI-k és a megtérülés rendszeres mérése a különböző forrásokból gyűjtött adatok - például az anyagfelhasználás, a munkaórák és a költségek - alapján lehetővé teszi a projektmenedzsment számára az erőforrások hatékony kezelését és a gyors döntések meghozatalát. Ezen adatok hosszú távú tárolása lehetővé teszi a jövőbeli tendenciák elemzését és a folyamatok optimalizálását.

A KPI-k, ROI-k és egyéb mérőszámok vizualizálására különböző grafikonokat és diagramok használnak, amelyeket általában műszerfalakba kombinálnak.

Műszerfalak és műszerfalak: a mérőszámok vizualizálása a hatékony irányítás érdekében

A mutatók és mérőszámok vizualizálására különféle diagramok és grafikonok szolgálnak, amelyeket általában adatbemutatókban és műszerfalakban egyesítnek. Ezek a műszerfalak központi áttekintést nyújtanak a projekt vagy a projekt egyes részeinek állapotáról, és (ideális esetben valós időben) megjelenítik a legfontosabb mutatókat. A naprakész és folyamatosan frissített műszerfalak lehetővé teszik a csapat számára, hogy gyorsan reagáljon a változásokra.

Az iránytáblák olyan eszközök, amelyek vizualizálják a mennyiségi értékeléseket, és így könnyen hozzáférhetővé és érthetővé teszik azokat a projekt minden résztvevője számára.



Ábra 7.1-6 A KPI-k kezelése és azok műszerfalakon keresztül történő megjelenítése kulcsfontosságú a projekt termelékenységének és hatékonyságának javításához.

Íme néhány példa a népszerű eszközökre, amelyekkel műszerfalakat hozhat létre:

- **A Power BI** a Microsoft eszköze interaktív jelentések és műszerfalak készítésére.
- **A Tableau és a Google Data Studio** hatékony eszközök az adatok vizualizálására és műszerfalak készítésére anélkül, hogy kódot kellene írni.
- **A Plotly** (7.1-6. ábra, 7.2-12. ábra) egy interaktív grafikonok készítésére szolgáló könyvtár, a Dash pedig egy keretrendszer az adatok elemzésére szolgáló webes alkalmazások létrehozására. Ezek együttesen interaktív műszerfalak létrehozására használhatók.
- **Számos Python könyvtár** (7.2-9. ábra - 7.2-11. ábra) - A Python számos nyílt forráskódú és ingyenes könyvtárral rendelkezik az adatvizualizációhoz, mint például a Matplotlib, Seaborn, Plotly, Bokeh és mások. Ezek segítségével grafikonok készíthetők, és integrálhatók webes alkalmazásokba olyan keretrendszerrel segítségével, mint a Flask vagy a Django.
- **JavaScript könyvtárak**: lehetővé teszi interaktív műszerfalak létrehozását nyílt forráskódú JavaScript könyvtárak, például a D3.js vagy a Chart.js használatával, és ezek integrálását egy webes alkalmazásba.

A KPI-k értékeléséhez és a műszerfalak létrehozásához naprakész adatokra és az információk összegyűjtésének és elemzésének világos ütemtervére van szükség.

Általában a KPI, a ROI és a műszerfalak az építőiparban a projektmenedzsment analitikus megközelítésének alapját képezik. Nemcsak az aktuális állapot nyomon követését és értékelését segítik, hanem értékes betekintést nyújtanak a jövőbeli tervezési és optimalizálási folyamatokhoz is - olyan folyamatokhoz, amelyek közvetlenül az adatok értelmezésétől és a megfelelő és időben feltett kérdésektől függnek.

Adatelemzés és a kérdésfeltevés művészete

Az adatok értelmezése az elemzés utolsó szakasza, amikor az információ értelmet nyer és "beszélni" kezd. Itt fogalmazódnak meg a kulcskérdezésekre adott válaszok: *"mit tegyünk?"* és *"hogyan tegyünk?"* (2.2-5. ábra). Ez a szakasz lehetővé teszi az eredmények összegzését, a minták azonosítását, az ok-okozati összefüggések megállapítását és a vizualizáció és statisztikai elemzésen alapuló következtetések levonását.

Talán nincs már messze az idő, amikor rájön, hogy ahoz, hogy a most kialakuló nagy, új, összetett világállapotok egyikének teljes mértékben hatékony polgárává váljon, éppúgy szükséges, hogy tudjon számolni, átlagokban, maximumokban és minimumokban gondolkodni, mint ahogyan ma már írni és olvasni is tudnia kell [133].

- Samuel S. Wilkes, az Amerikai Statisztikai Egyesület 1951-es elnöki beszédében idézett szavait idézi

Az Egyesült Királyság kormánya által közzétett "Adatelemzés és mesterséges intelligencia a kormányzati projektek megvalósításában" (2024) című jelentés [83] szerint az analitika adatok és a mesterséges intelligencia (AI) alkalmazása jelentősen javíthatja a projektmenedzsment-folyamatokat, növelheti az idő- és költségelőrejelzés pontosságát, valamint csökkentheti a kockázatot és a bizonytalanságot. A dokumentum kiemeli, hogy a fejlett analitikai eszközökkel használó állami szervezetek nagyobb teljesítményt érnek el az infrastrukturális kezdeményezésekben.

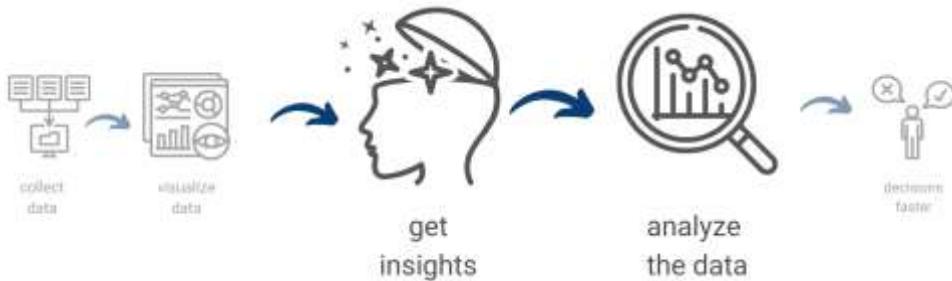
A negyedik ipari forradalom erős versenyhelyzetében és alacsony árréssel működő modern építőipari vállalkozások a katonai műveletekhez hasonlíthatók. Itt a vállalat túlélése és sikere az erőforrások és a minőségi információk beszerzésének gyorsaságától - és így az időben történő és megalapozott döntéshozataltól - függ (7.1-7. ábra).

Ha az adatvizualizáció az áttekintést biztosító "intelligencia", akkor az adatelemzés a cselekvéshez szükséges "muníció". Választ ad a kérdésekre: *mit* és *hogyan tegyünk*, és ez képezi az alapját a piaci versenyelőny megszerzésének.

Az analitika a széttagolt adatokat strukturált és értelmes információvá alakítja, amelyre a döntések alapozhatók.

Az elemzők és menedzserek feladata nem csupán az információk értelmezése, hanem megalapozott döntések meghozatala, a trendek azonosítása, a különböző típusú adatok közötti kapcsolatok meghatározása és kategorizálása a projekt céljainak és sajátosságainak megfelelően. A vizualizációs

eszközök és statisztikai elemzési módszerek segítségével az adatokat a vállalat stratégiai értékévé alakítják.



Ábra 7.1-7 Az adatelemzés az, amely az összegyűjtött információkat végső soron a döntéshozatal forrásává alakítja.

Ahhoz, hogy valóban megalapozott döntéseket hozhassunk az elemzési folyamat során, meg kell tanulni, hogyan kell helyesen megfogalmazni az adatokra vonatkozó kérdéseket. Ezeknek a kérdéseknek a minősége közvetlenül befolyásolja a nyert meglátások mélységét, és ennek következtében a vezetői döntések minőségét.

A múlt csak annyiban létezik, amennyiben jelen van a mai feljegyzésekben. Azt pedig, hogy ezek a feljegyzések mit képviselnek, az határozza meg, hogy milyen kérdéseket teszünk fel. Nincs más történelem, mint ez a történelem [134].

- John Archibald Wheeler, fizikus 1982

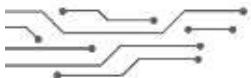
A mélyreható kérdések feltevésének és a kritikus gondolkodásnak a művészete kritikus készség az adatokkal való munka során. A legtöbb ember hajlamos egyszerű, felszínes kérdéseket feltenni, amelyek megválaszolása kevés erőfeszítést igényel. Az igazi elemzés azonban értelmes és átgondolt kérdésekkel kezdődik, amelyek képesek feltárnai a rejtett összefüggéseket és ok-okozati kapcsolatokat az információkban, amelyek az érvelés több rétege mögött rejtőzhetnek.

Az "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] című tanulmány szerint a sikeres digitális átalakuláshoz elemzési képességekbe, változáskezelési programokba, valamint az üzleti célok és az IT-kezdeményezések összehangolásába való befektetésre van szükség. Az adatvezérelt kultúrát kialakító vállalatoknak be kell fektetniük az adatalemzési képességekbe, és változáskezelési programokat kell indítaniuk az új gondolkodásmód, viselkedésmódok és munkamódszerek meghonosítása érdekében.

Az elemzési kultúra fejlesztésére, az adatgyűjtési eszközök javítására és a szakemberek képzésére irányuló beruházások nélkül a vállalatok továbbra is azt kockáztatják, hogy elavult vagy hiányos információkon alapuló döntéseket hoznak - vagy a HiPPO vezetőinek szubjektív véleményére hagyatkoznak (2.1-9. ábra).

Az analitika és a műszerfalak relevanciájának és folyamatos frissítésének szükségességének felismerése elkerülhetetlenül arra készti a vezetést, hogy megértse az analitikai folyamatok automatizálásának fontosságát. Az automatizálás növeli a döntéshozatal sebességét, csökkenti az emberi tényező hatását és biztosítja az adatok relevanciáját. Az információmennyiségek exponenciális növekedésével a sebesség nem csupán versenyelőnnyé, hanem a fenntartható siker kulcstényezőjévé válik.

Az adatelemzési és -feldolgozási folyamatok automatizálása általában elválaszthatatlanul kapcsolódik az ETL (Extract, Transform, Load) témaréhez. Ahogyan az automatizálási folyamat során az adatokat át kell alakítani, úgy az ETL-folyamat során az adatokat különböző forrásokból vonjuk ki, a szükséges követelményeknek megfelelően alakítjuk át, és töltjük be a célrendszerbe további felhasználás céljából.



FEJEZET 7.2.

ADATÁRAMLÁS KÉZI MUNKA NÉLKÜL: MIÉRT VAN SZÜKSÉG AZ ETL-RE

ETL automatizálás: költségcsökkentés és gyorsabb adatkezelés

Amikor a kulcsfontosságú teljesítménymutatók (KPI-k) az adatmennyiség és a csapat méretének növekedése ellenére nem növekednek tovább, a vállalat vezetése elkerülhetetlenül rájön, hogy szükség van a folyamatok automatizálására. Ez a felismerés előbb-utóbb ösztönzőleg hat a komplex automatizálás elindítására, amelynek fő célja a folyamatok összetettségének csökkentése, a feldolgozás felgyorsítása és az emberi tényezőtől való függőség csökkentése.

A McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] című tanulmánya szerint a streaming adatarchitektúrákat használó vállalatok jelentős előnyre tesznek szert, mivel valós időben tudják elemezni az információkat. A streaming technológiák lehetővé teszik a valós idejű üzenetek közvetlen elemzését és a gyártásban a prediktív karbantartás alkalmazását a valós idejű szenzoradatok elemzésével.

A folyamatok egyszerűsítése az automatizálás, ahol a hagyományos kézi funkciókat algoritmusok és rendszerek váltják fel.

Az automatizálás, vagy inkább "az ember szerepének minimalizálása az adatfeldolgozásban" minden vállalat számára visszafordíthatatlan és rendkívül érzékeny folyamat. Bármely szakmai terület szakemberei gyakran vonakodnak teljesen feltárnivaló módszereiket és munkájuk finomságait az optimalizáló kollégák előtt, mivel tisztában vannak azzal a kockázattal, hogy a gyorsan fejlődő technológiai környezetben elveszíthetik állásukat.

Ha ellenségeket akarsz szerezni, próbálj meg változtatni a dolgokon [137].

- Woodrow Wilson, beszéd a kereskedők kongresszusa előtt, Detroit, 1916

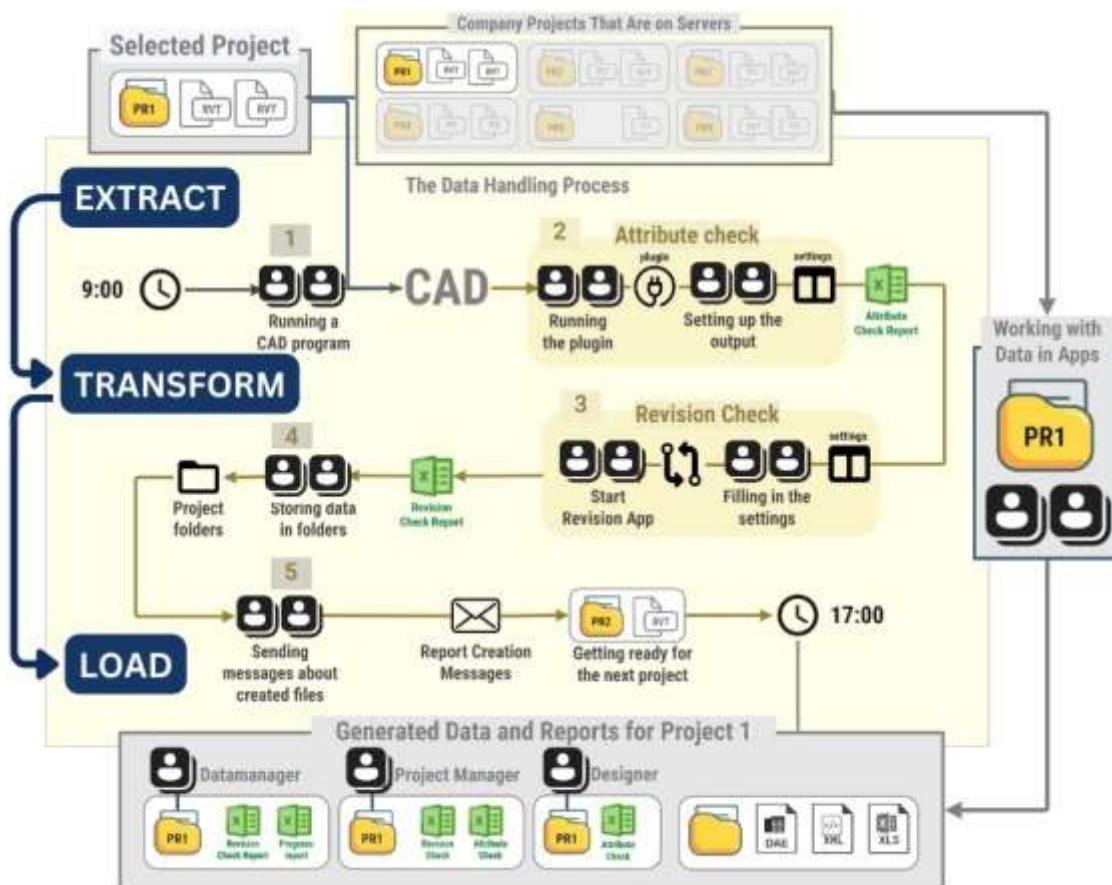
Az automatizálás nyilvánvaló előnyei ellenére sok vállalat napi gyakorlatában még mindig nagy a kézi munka aránya, különösen a mérnöki adatok területén. A jelenlegi helyzet illusztrálására nézzünk meg egy tipikus példát az ilyen folyamatokon belüli szekvenciális adatfeldolgozásra.

A kézi adatkezelést a CAD-adatbázisokból származó információkkal való interakció példájával lehet szemléltetni. A hagyományos adatfeldolgozás ("kézi" ETL-folyamat) a CAD (BIM) részlegekben az attribútumtáblák létrehozására vagy a tervezési adatokon alapuló dokumentáció elkészítésére a következő sorrendben zajlik (7.2-1. ábra):

1. **Kézi kivonás (Extract):** a felhasználó manuálisan megnyitja a projektet - a CAD-alkalmazás (BIM) elindításával (7.2-1 ábra 1. lépés).
2. **Ellenőrzés:** a következő lépés általában több plug-in vagy a segédalkalmazás kézi futtatását jelenti az adatok előkészítése és minőségének értékelése céljából (7.2-1 2-3. ábra).
3. **Kézi átalakítás (Transform):** az előkészítés után kezdődik az adatfeldolgozás, amely különböző

szoftvereszközök kézi működtetését igényli, amelyekben az adatokat előkészítik a feltöltéshez (7.2-1. ábra 4. lépés).

4. Kézi feltöltés (Load): a konvertált adatok kézi feltöltése külső rendszerekbe, adatformátumokba és dokumentumokba (7.2-1. ábra 5. lépés).



ábra 7.2-1 A hagyományos kézi ETL feldolgozást az egyes technikusok vágyai és fizikai képességei korlátozzák.

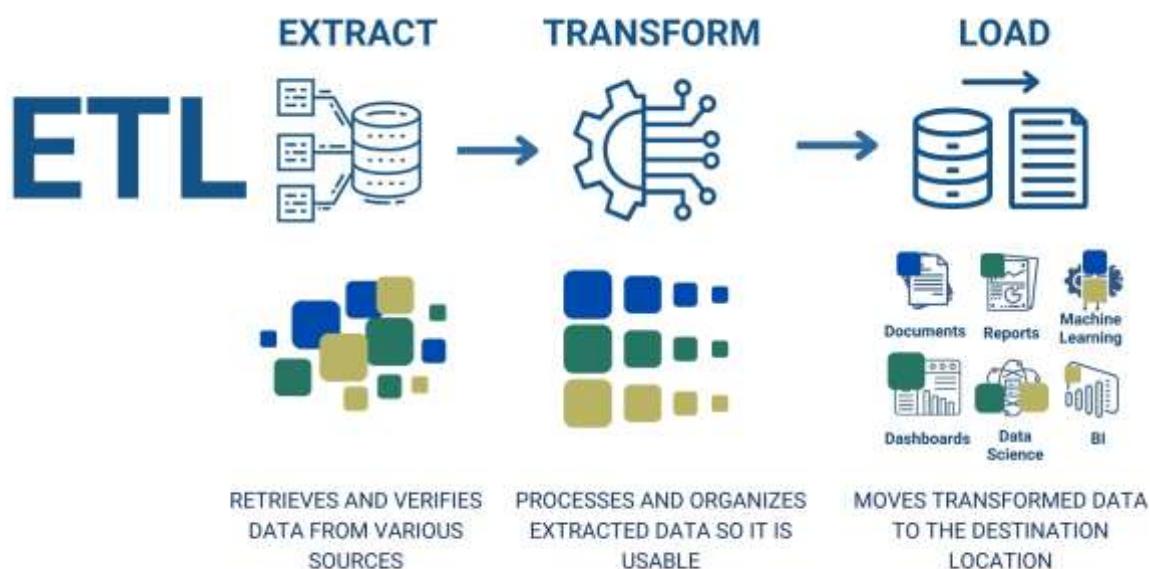
Egy ilyen munkafolyamat egy klasszikus ETL-folyamat - extrakció, transzformáció és betöltés (ETL) példája. Más iparágakkal ellentétben, ahol az automatikus ETL-csővezetékek már régóta szabványnak számítanak, az építőiparban még mindig a kézi munka dominál, ami lassítja a folyamatokat és növeli a költségeket

Az ETL (Extract, Transform, Load) az adatok különböző forrásokból történő kinyerése, a kívánt formátumba való átalakítása és a célrendszerbe való betöltése a további elemzés és felhasználás céljából.

Az ETL egy olyan folyamat, amely az adatfeldolgozás három fő összetevőjét jelöli: Extract, Transform és Load (7.2-2. ábra):

- Extract - adatok kinyerése különböző forrásokból (fájlok, adatbázisok, API).
- **Transform** - adattisztítás, aggregálás, normalizálás és logikai feldolgozás.
- **Betöltés** - strukturált információk betöltése adattárházba, jelentésbe vagy BI-rendszerbe.

A könyv korábbi részeiben az ETL fogalmát csak alkalmanként érintettük: egy strukturálatlan szkennelt dokumentum strukturált táblázatos formátumba való átalakításakor (4.1-1. ábra), a követelmények formalizálásának összefüggésében, hogy rendszerezzük mind az élet-, mind az üzleti folyamatok érzékelését (4.4-20. ábra), valamint az adatellenőrzés és a CAD-megoldásokból történő adatfeldolgozás automatizálása során. Tekintsük most részletesebben az ETL-t a tipikus munkafolyamatok összefüggésében.



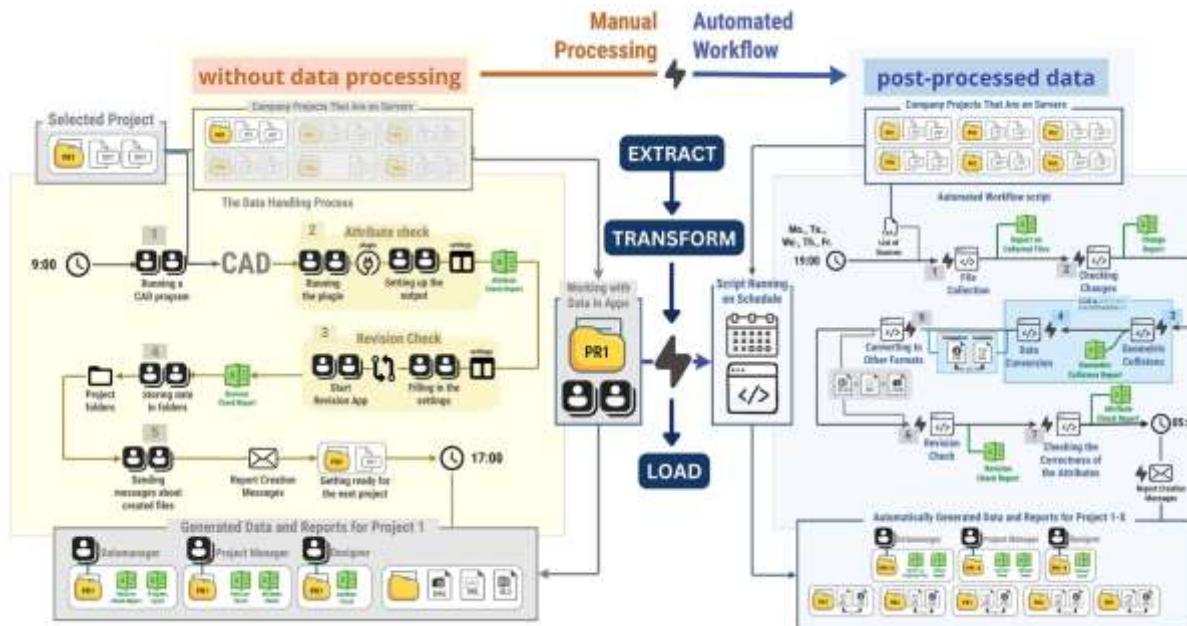
7.2-2. ábra ETL automatizálja az ismétlődő adatfeldolgozási feladatokat.

Kézi vagy félautomata ETL -folyamat olyan menedzsert vagy technikust feltételez, aki minden lépést kézzel kezel - az adatgyűjtéstől a jelentéskészítésig. Egy ilyen folyamat jelentős időt vesz igénybe, különösen, ha a munkaidő korlátozott (pl. 9:00-17:00 óra).

A vállalatok az alacsony hatékonyság és a lassúság problémáját gyakran moduláris integrált megoldások (ERP, PMIS, CPM, CAFM, stb.) megvásárlásával próbálják megoldani, amelyeket aztán külső gyártók és tanácsadók fejlesztenek tovább. Ezek a szállítók és külső fejlesztők azonban gyakran kritikus függőségi ponttá válnak: technikai korlátaik közvetlenül befolyásolják az egész rendszer és a vállalkozás egészének teljesítményét, amint azt a saját rendszerekről és formátumokról szóló korábbi fejezetekben részleteztük. A széttagoltság és a függőség által okozott problémákat részletesen tárgyalta a "Hogyan fulladtak bele az építőipari vállalkozások az adatkáoszba" című fejezetben.

Ha egy vállalat nem áll készen arra, hogy az egyik szállítótól származó nagy moduláris platformot valósítson meg, akkor alternatív automatizálási módok után kezd el kutatni. Az egyik ilyen lehetőség a saját moduláris, nyílt ETL-konzerverek kifejlesztése, ahol minden egyes fázis (extrakció, transzformáció, validálás, betöltés) ütemezetten végrehajtott szkriptek formájában valósul meg.

Ugyanezen ETL-munkafolyamat automatizált változatában (7.2-1. ábra) a munkafolyamat úgy néz ki, mint egy moduláris kód, amely az adatok feldolgozásával és nyílt, strukturált formába való lefordításával kezdődik. A strukturált adatok beérkezése után a különböző szkriptek vagy modulok automatikusan, ütemezetten futnak a változások ellenőrzése, átalakítása és üzenetek küldése céljából (7.2-3. ábra).



7.2-3 A bal oldalon a kézi megmunkálás, a jobb oldalon pedig egy olyan automatikus folyamat látható, amelyet a hagyományos kézi megmunkálással ellentétben nem korlátoznak a felhasználó képességei.

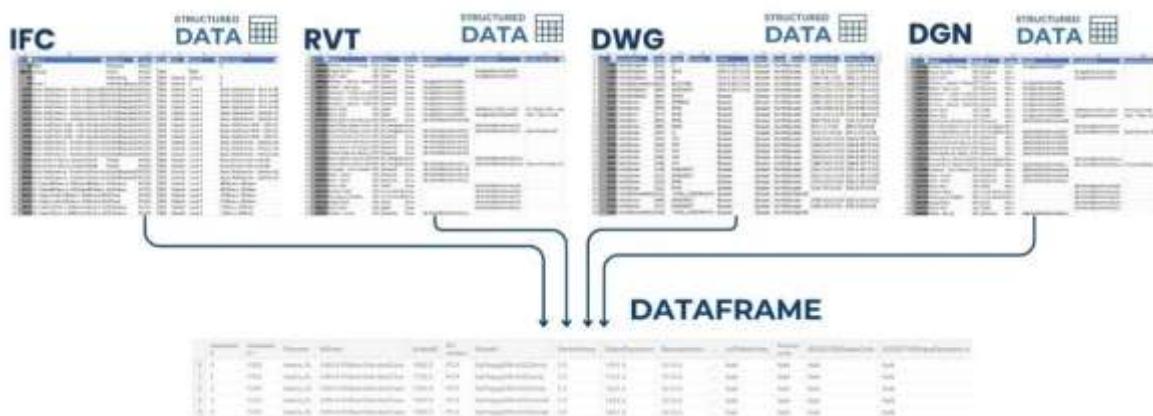
Az automatizált munkafolyamatban az adatfeldolgozást egyszerűsíti az ET(L) adatelőfeldolgozás: strukturálás és egységesítés.

A hagyományos feldolgozási módszerekben a szakemberek az adatokkal úgy dolgoznak, ahogyan azok vannak - ahogyan azokat a rendszerekből vagy szoftverekből lekérlik. Ezzel szemben az automatizált folyamatok során az adatok gyakran először egy ETL-pipeline-on haladnak át, ahol a további felhasználásra és elemzésre alkalmas, konzisztens szerkezetbe és formátumba kerülnek.

Vegyük egy gyakorlati ETL-példát, amely a "adatok és az érvényesítési eredmények hitelesítése" című fejezetben leírt adattábla-érvényesítési folyamatot mutatja be (4.4-13. ábra). Ehhez a Pandas könyvtárat az LLM-mel együtt használjuk az automatizált adatelemzési és adatfeldolgozási folyamatokhoz.

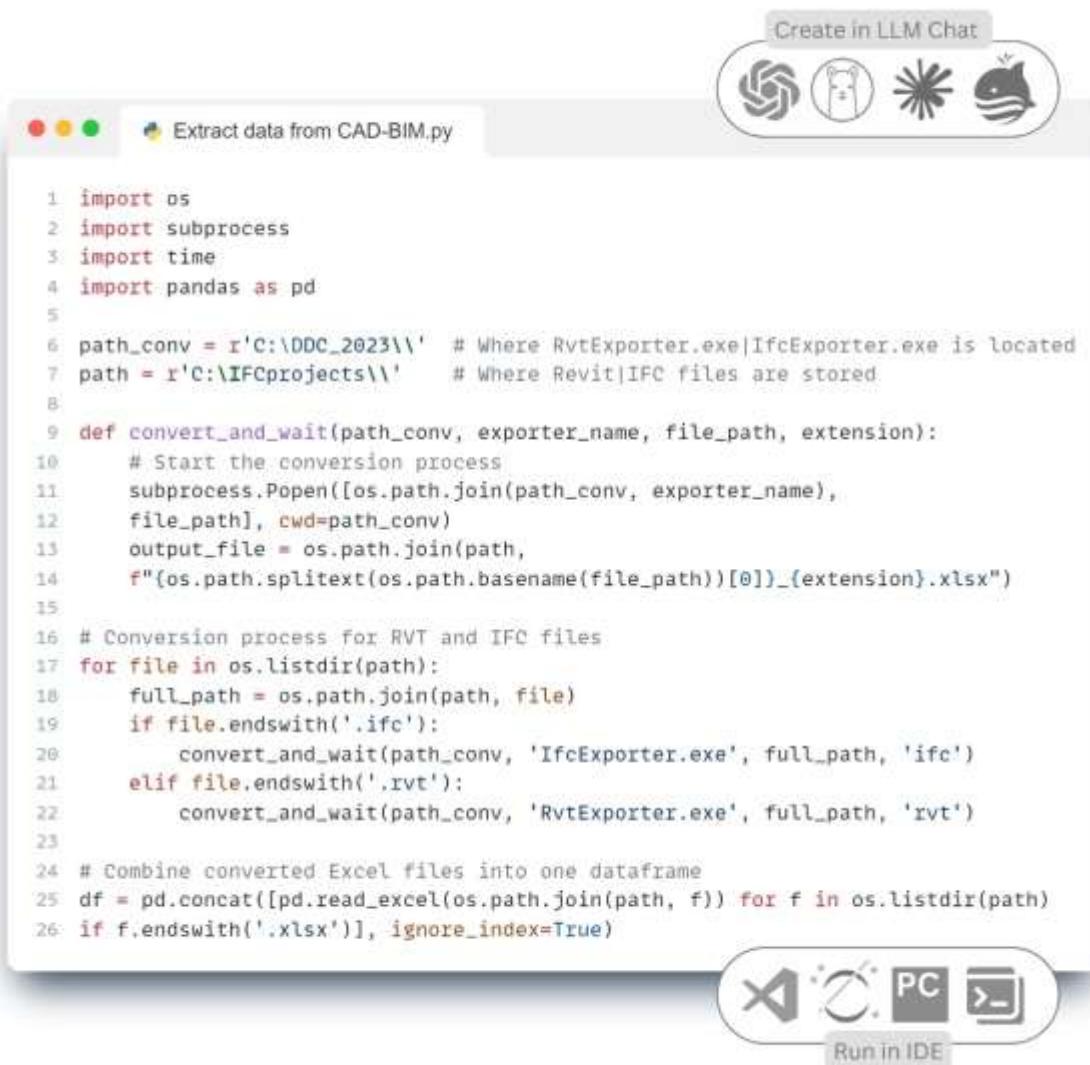
ETL Kivonat: adatgyűjtés

Az ETL folyamat első szakasza - Extract) - a további ellenőrzésre és feldolgozásra szánt adathalmazok összegyűjtésére szolgáló kód megírásával kezdődik. Ehhez beolvassuk a termelő szerver összes mappáját, összegyűjtjük a meghatározott formátumú és tartalmú dokumentumokat, majd strukturált formába alakítjuk őket. Ezt a folyamatot részletesen a "Strukturálatlan és szöveges adatok strukturált formába történő átalakítása" és a "CAD-adatok (BIM) strukturált formába történő átalakítása" című fejezetek tárgyalják (4.1-1. ábra - 4.1-12. ábra).



Ábra 7.2-4 Konvertálja a CAD-adatokat (BIM) egyetlen nagy adatkeretbe, amely tartalmazza az összes projektrészletet.

Szemléltető példaként az Extract adatbetöltési lépést használjuk, és az összes CAD- (BIM-) projektet tartalmazó táblázatot. 2-4. ábra a reverse engineeringet lehetővé tevő [138] RVT és IFC formátumú konvertereit használja az összes projekt strukturált táblázatainak megszerzéséhez, és egyetlen nagy DataFrame táblában egyesíti őket.



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\\' # Where RvtExporter.exe|IfcExporter.exe is located
7 path = r'C:\IFCprojects\\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10     # Start the conversion process
11     subprocess.Popen([os.path.join(path_conv, exporter_name),
12                      file_path], cwd=path_conv)
13     output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16 # Conversion process for RVT and IFC files
17 for file in os.listdir(path):
18     full_path = os.path.join(path, file)
19     if file.endswith('.ifc'):
20         convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21     elif file.endswith('.rvt'):
22         convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24 # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

[Run in IDE](#)

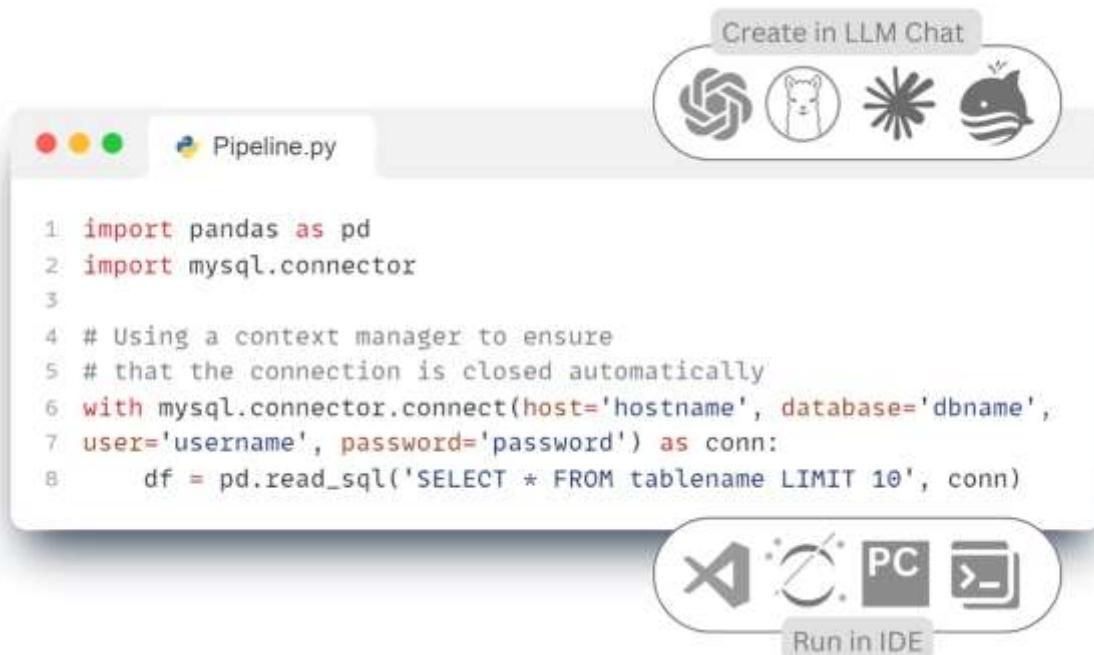
Ábra 7.2-5 Python kód és SDK reverse engineering eszköz segítségével történő átalakítás az RVT és az IFC fájlokat egy nagy strukturált (df) DataFrame.

A Pandas DataFrame képes adatokat betölteni különböző forrásokból, beleértve a CSV szöveges fájlokat, Excel táblázatok, JSON - és XML - fájlokat, nagy adattárolási formátumokat, mint a Parquet és HDF5, valamint a MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle és más adatbázisokból. Ezenkívül a Pandas támogatja az adatok betöltését API-kból, weboldalakról, felhőszolgáltatásokból és tárolórendszerkből, mint például a Google BigQuery, az Amazon Redshift és a Snowflake.

- ❷ Ha kódot szeretne írni az adatbázisokhoz való csatlakozáshoz és az onnan történő információgyűjtéshez, küldjön egy hasonló szöveges kérést az LLM csevegőszobába (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vagy bármely más):

Kérjük, írjon egy példát a MySQL-hez való csatlakozásra és az adatok való konvertálására. ↵

💡 LLM válasza :



```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7     user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

7.2-6 ábra Példa a Python segítségével történő csatlakozásra egy MySQL adatbázishoz és az adatok importálása a MySQL adatbázisból egy DataFrame-be .

Az így kapott kód (7.2-5. ábra, 7. ábra).2-6) futtatható a fent említett népszerű IDE-k (integrált fejlesztőkörnyezetek) valamelyikében offline módban: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse PyDev pluginnal, Thonny, Wing IDE, IntelliJ IDEA Python pluginnal, JupyterLab vagy népszerű online eszközök: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

A multiformátumú adatokat a "df" változóba betölvtve (7.2-5. ábra - 25. sor; 7.2-6. ábra - 8. sor), az adatokat a Pandas DataFrame formátumba konvertáltuk, amely az adatfeldolgozás egyik legnépszerűbb struktúrája, és amely egy kétdimenziós táblázat sorokkal és oszlopokkal. Az ETL -Pipeline-okban használt egyéb tárolási formátumokról, mint például a Parquet, Apache ORC, JSON, Feather, HDF5, valamint a modern adattárházakról az "Adattárolás és -kezelés az építőiparban" című fejezetben (8.1-2. ábra) bővebben beszélünk.

Az adatok kinyerésének és strukturálásának szakasza (Extract) után egyetlen információs tömb jön létre (7.2-5. ábra, 7.2-6. ábra), amely készen áll a további feldolgozásra. Mielőtt azonban ezeket az adatokat betöltenénk a célrendszerbe vagy felhasználnánk elemzésre, biztosítani kell a minőségüket, integritásukat és a meghatározott követelményeknek való megfelelésüket. Ebben a szakaszban kerül sor az adatok átalakítására (Transform), ami kulcsfontosságú lépés a későbbi következtetések és döntések megbízhatóságának biztosításához.

ETL Transform: validálási és transzformációs szabályok alkalmazása

A Transform lépés az adatok feldolgozása és átalakítása. Ez a folyamat magában foglalhatja a helyesség ellenőrzését, a normalizálást, a hiányzó értékek kitöltését és az automatizált eszközökkel történő validálást

A PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] című tanulmány szerint a modern könyvvizsgáló cégek eltávolodnak a véletlenszerű adatellenőrzéstől, és a hatalmas mennyiségű információ automatizált eszközökkel történő elemzése felé mozdulnak el. Ez a megközelítés lehetővé teszi nemcsak a beszámolókban lévő eltérések azonosítását, hanem azt is, hogy ajánlásokat tegyenek az üzleti folyamatok optimalizálására.

Az építőiparban hasonló módszerek alkalmazhatók például a tervezési adatok automatikus validálására, az építési minőségellenőrzésre és a vállalkozói teljesítmény értékelésére. Az adatfeldolgozás automatizálásának és felgyorsításának egyik eszköze a reguláris kifejezések (RegEx) használata az ETL-folyamat adattranszformációs fázisában (Transform). A RegEx lehetővé teszi az adatsorok hatékony érvényesítését, az ellentmondások felderítését és az információk integritásának biztosítását minimális erőforrás-felhasználás mellett. A RegEx (4.4-7. ábra) további részleteit a "Követelmények strukturált formába történő lefordítása" című fejezetben tárgyaltuk.

Nézzünk egy gyakorlati példát: a tulajdonságobjektum-kezelő rendszerben (RPM) a menedzser meghatározza a követelményeket az objektumok kulcsfontosságú attribútumaihoz (7.2-7. ábra). Az átalakítási szakaszban a következő paramétereket kell érvényesíteni:

- az objektumazonosító formátumok ellenőrzése ("ID" attribútum)
- a jótállási időre vonatkozó csereértékek ellenőrzése ("jótállási idő" attribútum)
- az elemcsere-ciklus ellenőrzése ("Kabantartási követelmények" attribútum)



The screenshot shows the 'Property Manager: Long-term Management' section of the RPM software. On the left, there's a logo for 'Property Manager' and 'Real Property Management' next to a database icon. The main area displays a table with the following data:

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

7.2-7. ábra Ellenőrzés A minőség az attribútumokra és azok határértékeire vonatkozó követelmények meghatározásával kezdődik.

A paraméterek érvényesítésére vonatkozó határértékek beállításához például tegyük fel, hogy a oldalról tudjuk, hogy az "ID" attribútum érvényes értékei csak a "W-NEW", "W-OLD1" vagy "D-122" vagy hasonló értékek lehetnek, ahol az első karakter egy betű, amelyet egy kötőjel követ, majd három betűvel: "NEW", "OLD"

vagy bármely háromjegyű szám (7.2-7. ábra). A következő reguláris kifejezés (RegEx) használható ezen azonosítók érvényesítésére:

```
^W-NEW$ | ^W-OLD[0-9]+$ | ^D-1[0-9]{2}$
```

Ez a sablon biztosítja, hogy az adatokban szereplő összes azonosító megfeleljen a megadott kritériumoknak. Ha valamelyik érték nem felel meg az ellenőrzésnek, a rendszer hibát ír ki. Az adatok átalakítására szolgáló Python -kód létrehozásához és az így kapott adatok felhasználásához egy eredménytábla létrehozásához egyszerűen fogalmazzon meg egy lekérdezést az LLM chatban.

💡 Szöveges kérelem az LLM címre:

Írjon kódot a DataFrame oszlopainak ellenőrzésére a, amely a RegEx segítségével ellenőrzi a "W-NEW" vagy "W-OLD" formátumú azonosítókat, az "A"-tól "G"-ig terjedő betűjelű energiahatékonyságot, a garanciális időszakot és a csereciklust numerikus értékekkel, években kifejezve. ↴

💡 LLM válasza :

```
1 # Regex
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2}$',
7     'Energy_Efficiency': r'^[A-G]\w*$',
8     'Warranty_Period': r'^\d+\w+years$',
9     'Replacement_Cycle': r'^\d+\w+years$'
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20
```

7.2-8. ábra A kód automatizálja az érvényesítési folyamatot a RegEx sablonok adatkeret paraméter oszlopokra történő alkalmazásával .

A fenti, automatikusan generált Python -kód (7.2-8. ábra) a "re" könyvtárat (RegEx regular expressions) használja egy olyan függvény definíálására, amely egy DataFrame adatelem minden egyes attribútumát

ellenőrzi. A függvény minden egyes megadott oszlopra (attribútumra) egy RegEx mintát alkalmaz annak ellenőrzésére, hogy az egyes rekordok megfelelnek-e az elvárt formátumnak, és az eredményeket új értékként (False/True) hozzáadja a DataFrame új attribútum-oszlopához.

Az ilyen automatizált validálás biztosítja az adatok formális megfelelőségét, és az átalakítási fázisban a minőségellenőrzési rendszer részeként használható.

A Transform lépés és a minőségellenőrzés sikeres elvégzése után az adatok feltölthetők a célrendszerekbe. A transzformált és validált adatok további felhasználás céljából feltölthetők CSV, JSON, Excel, adatbázisokba és más formátumba. A feladattól függően az eredmények jelentésekben, grafikonokon vagy analitikus műszerfalakon is bemutathatók.

ETL Load: A eredmények grafikonokon és diagramokon való megjelenítése.

A Transform szakasz befejezése után, amikor az adatokat strukturálták és ellenőrizték, az utolsó szakasz a Load, ahol az adatokat be lehet tölteni a célrendszerbe, és az elemzéshez megjeleníthetők. Az adatok vizuális megjelenítése lehetővé teszi az eltérések gyors azonosítását, az eloszlások elemzését és a legfontosabb következtetések közlését a projekt valamennyi résztvevőjével, beleértve a technikai háttérrel nem rendelkezőket is.

Az információk táblázatok és számok formájában történő bemutatása helyett használhatunk infografikákat, grafikonokat és műszerfalakat (dashboardokat). Az egyik legelterjedtebb és legrugalmasabb eszköz a strukturált adatok Pythonban történő megjelenítésére a Matplotlib könyvtár (7.2-9. ábra, 7.2-10. ábra). Lehetővé teszi statikus, animált és interaktív diagramok készítését, és a diagramtípusok széles skáláját támogatja.

- Az RPM-rendszer attribútum-ellenőrzésének eredményeinek megjelenítéséhez (7.2-7. ábra), a következő lekérdezéssel lehet a nyelvi modellhez:

Írjon kódot a fenti DataFrame adatainak vizualizálásához (7.2-7. ábra), az eredmények hisztogramjával, amely megmutatja az attribútum hibáinak gyakoriságát. ↵

- ❸ LLM válasz kód formájában és a kódvégrehajtás eredményeinek kész megjelenítése közvetlenül az LLM chatszobában :

The screenshot shows a chat interface with a Python script named "Pipeline.py". The code imports matplotlib.pyplot and seaborn, then creates a stacked bar chart for data verification. The chart has four categories on the x-axis: Window_ID, Energy_Efficiency, Warranty_Period, and Replacement_Cycle. Each category has two segments: a green segment at the bottom labeled "Passed" and a red segment above it labeled "Failed". The y-axis is labeled "Count (of values checked)" and ranges from 0 to 5.

```

1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()

```

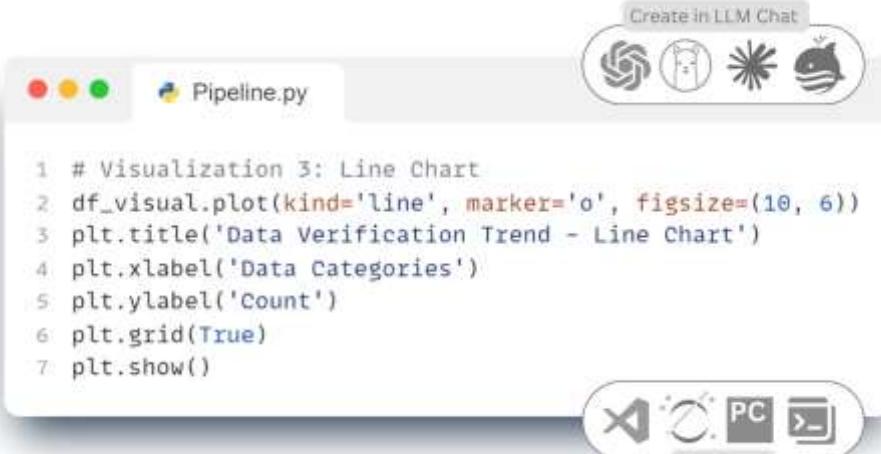
Data Category	Passed (Green)	Failed (Red)	Total Count
Window_ID	4	1	5
Energy_Efficiency	3	2	5
Warranty_Period	4	1	5
Replacement_Cycle	3	2	5

7.2-9. ábra A ábra A lépés eredményeinek vizualizálása a transzformáció az RPM rendszerből származó attribútumértékek ellenőrzésének (7.2-7. ábra) hisztogram formájában a Load lépében.

- Számos nyílt forráskódú és ingyenes vizualizációs könyvtár létezik, amelyek lehetővé teszik a strukturált adatok különböző formátumokban történő megjelenítését. Folytassuk az eredmények vizualizálását egy másik típusú grafikonnal, a következő kéressel a csevegőben:

Ugyanezen adatok ábrázolása egyenesek grafikonjaként ↗

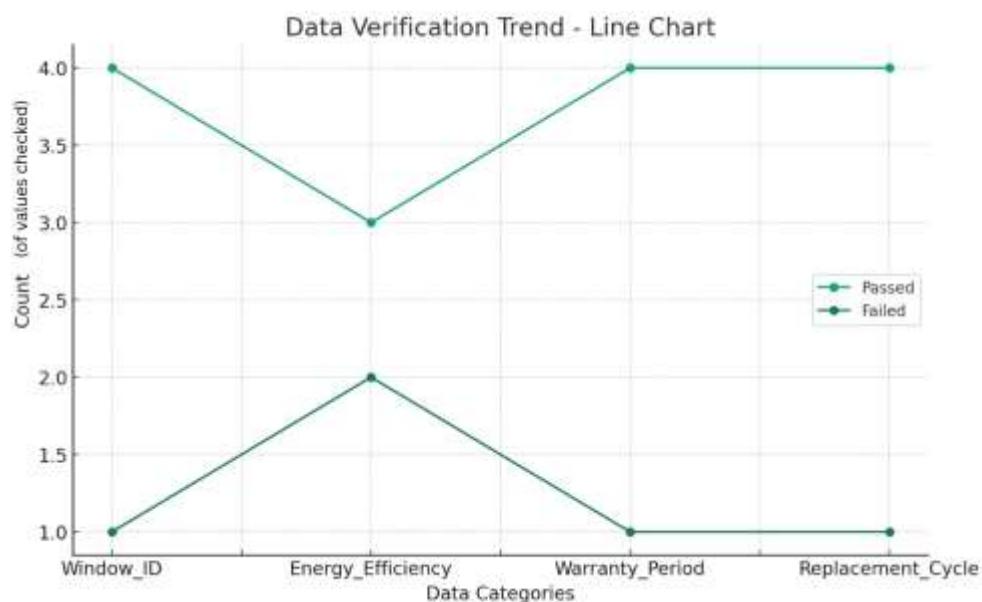
- ❷ LLM válasza :



```

1 # Visualization 3: Line Chart
2 df_visual.plot(kind='line', marker='o', figsize=(10, 6))
3 plt.title('Data Verification Trend - Line Chart')
4 plt.xlabel('Data Categories')
5 plt.ylabel('Count')
6 plt.grid(True)
7 plt.show()

```



7.2-10. ábra A validálási adatok (7.2-8. ábra) vizualizációja a Matplotlib könyvtár segítségével készült vonaldiagram formájában.

Számos nyílt forráskódú és ingyenes vizualizációs könyvtár létezik, mint például:

- Seaborn - statisztikai grafikonokhoz (7.2-11. ábra)
- Plotly - interaktív webes megjelenítésekhez (7.2-12 ábra, 7.1-6 ábra).
- Altair - a deklaratív vizualizációhoz
- Dash vagy Streamlit - teljes műszerfalak létrehozásához

A vizualizáláshoz nem szükséges speciális könyvtárak ismerete - a modern eszközök, köztük az LLM, lehetővé teszik, hogy a feladatleírás alapján automatikusan generálja a grafikonok és teljes alkalmazások kódját.

Az eszköz kiválasztása a projekt célkitűzéseitől függ: jelentés, prezentáció vagy online műszerfal. A Seaborn nyílt forráskódú könyvtár például különösen jól használható kategorikus adatokkal való munkához, segít a minták és trendek azonosításában.

- ▣ A Seaborn könyvtár működés közbeni megtekintéséhez a címen vagy közvetlenül az LLM-től kérheti a kívánt könyvtár használatát, vagy küldhet egy hasonló szöveges kérést az LLM-nek a címre:

Hőterkép megjelenítése az ↵

- ▣ LLM válasza kód és egy kész grafikon formájában, amelynek ábrázoló kódja most már bemásolható az IDE-be, maga a grafikon pedig másolható vagy menthető egy dokumentumba való beillesztéshez:

```

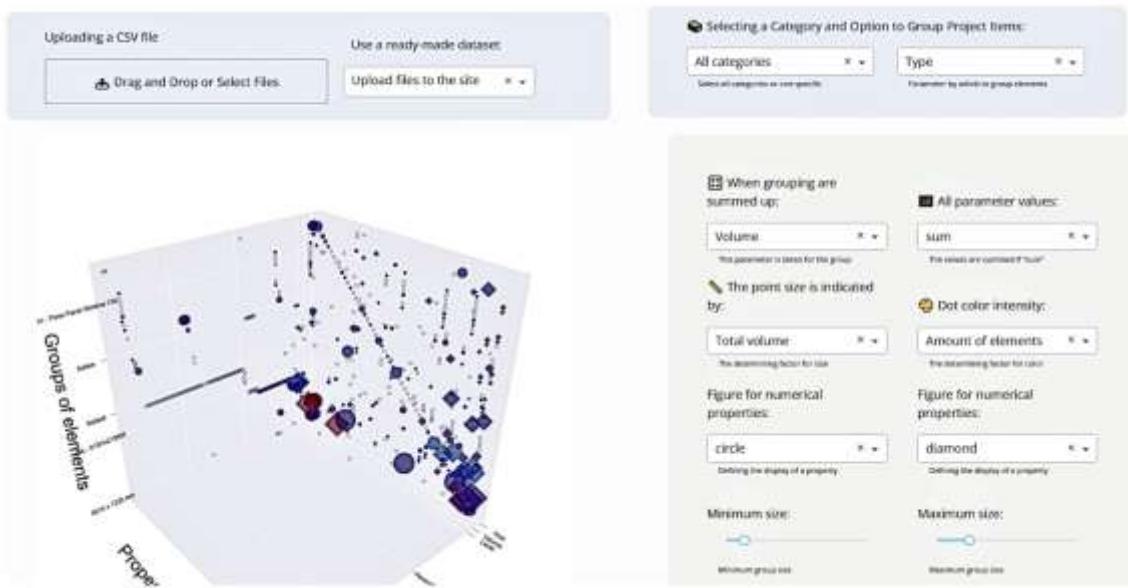
1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

```



7.2-11. ábra A vizualizálás az adatok validálásának (7.2-8. ábra) eredményei a Seaborn könyvtár segítségével.

Azok számára, akik az interaktív megközelítést részesítik előnyben, léteznek olyan eszközök, amelyekkel dinamikus diagramokat és paneleket hozhat létre interakcióképességgel. A Plotly könyvtár (7.1-6. ábra, 7.2-12. ábra) lehetőséget kínál rendkívül interaktív diagramok és panelek létrehozására, amelyek weboldalakba ágyazhatók, és lehetővé teszik a felhasználó számára, hogy valós időben interakcióba lépjen az adatokkal.



7.2-12 ábra Interaktív 3D -a CAD- (BIM-) projektből származó elemattribútumok interaktív 3D - visualizációja a Plotly könyvtár segítségével.

A Bokeh, a Dash és a Streamlit speciális nyílt forráskódú könyvtárak kényelmes módot biztosítanak az adatok megjelenítésére anélkül, hogy mély webfejlesztői ismeretekre lenne szükség. A Bokeh összetett interaktív grafikonok készítésére alkalmas, a Dash teljes értékű analitikus műszerfalak készítésére, a Streamlit pedig adatelemzésre szolgáló webes alkalmazások gyors létrehozását teszi lehetővé.

Az ilyen vizualizációs eszközökkel a fejlesztők és az elemzők hatékonyan terjeszthetik az eredményeket a kollégák és az érdekelt felek számára, lehetővé téve az adatokkal való intuitív interakciót és egyszerűsítve a döntéshozatalt.

ETL Load: PDF dokumentumok automatikus létrehozása.

Az adatbetöltési szakaszban nemcsak az adatok vizualizálása, táblázatokba vagy adatbázisokba való feltöltése lehetséges, hanem automatikusan készíthet jelentéseket is, beleértve a szükséges grafikonokat, diagramok és kulcsfontosságú elemzési mutatókat, amelyeket az ellenőrzés eredményeit váró vezető vagy szakember megkaphat. Az automatizált jelentések tartalmazhatnak megjegyzéseket és az adatok szöveges értelmezését, valamint vizualizációs elemeket - táblázatokat, grafikonokat - is.

- Egy PDF-jelentés készítéséhez, amely tartalmaz egy hisztogramot (7.2-9. ábra) és az előző fejezetekben elvégzett validáláson alapuló elemzés leírását, elegendő egy kérést megfogalmazni az LLM-mel folytatott párbeszéd folytatásában, pl.:

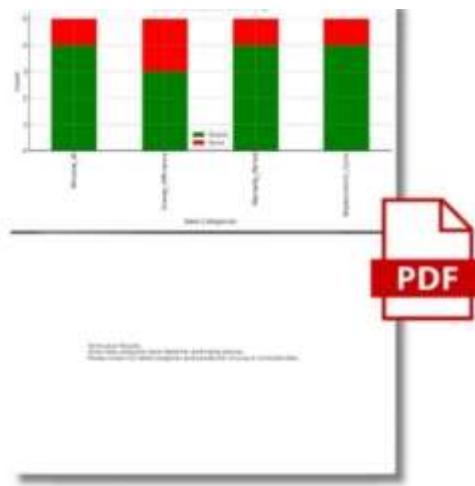
Írjon kódot egy PDF-fájl létrehozásához, amely tartalmaz egy hisztogramot és a fenti adatellenőrzés eredményeinek leírását (a chatben), és írjon egy szöveges figyelmeztetést, hogy néhány kategóriát nem érvényesítettek, és hogy a hiányzó ki kell tölteni. ↵

- LLM válasza kódként és kész PDF :

```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df.pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



7.2-13 ábra Az automatizált kód egy PDF -dokumentumot hoz létre, amely egy hiszogramot tartalmaz a tesztadatokkal és egy szöveget a teszteredményekkel.

A mindössze 20 sornyi kódot tartalmazó, az LLM segítségével automatikusan megírt megoldás azonnal létrehozza a kívánt PDF (vagy DOC) dokumentumot, amely egy attribútum-hiszogram formájában megjelenített vizualizációval rendelkezik (7.2-13. ábra), amely megmutatja a validáláson átment és azon megbukott adatok számát, valamint egy szövegblockk alapján kiegészítve, amely összefoglalja az eredményeket és a további lépésekre ajánlásokat.

Az automatizált dokumentumgenerálás a weboldalon a betöltési szakasz kulcseleme, különösen olyan projektkörnyezetben, ahol a jelentéskészítés gyorsasága és pontossága kritikus fontosságú.

ETL Load: dokumentumok automatikus generálása FPDF-ből

A jelentéskészítés automatizálása az ETL szakaszban Load az adatfeldolgozás fontos lépése, különösen akkor, ha az elemzések eredményeit könnyen kommunikálható és érhető formában kell bemutatni. Az építőiparban ez gyakran az előrehaladási jelentések, projektadat-statisztikák, minőségbiztosítási jelentések vagy pénzügyi dokumentációk esetében releváns.

Az egyik legkényelmesebb eszköz az ilyen feladatokhoz a, FPDF nyílt forráskódú könyvtár, amely Python és PHP nyelven is elérhető.

A nyílt forráskódú **FPDF** könyvtár rugalmasan használható a dokumentumok kódon keresztüli generálásához, lehetővé téve fejlécök, szöveg, táblázatok és képek hozzáadását. A kód használata a kézi szerkesztés helyett csökkenti a hibák számát és felgyorsítja a PDF formátumú jelentések elkészítésének folyamatát.

A PDF -dokumentum létrehozásának egyik legfontosabb lépése a címsorok és a fő szöveg hozzáadása megjegyzések vagy leírások formájában. Egy jelentés létrehozásakor azonban nemcsak a szöveg hozzáadása, hanem annak megfelelő strukturálása is fontos. Címsorok, behúzások, sortávolság - mindenek befolyásolják a dokumentum olvashatóságát. Az FPDF segítségével formázási paramétereket állíthat be, ellenőrizheti az elemek elrendezését és testre szabhatja a dokumentum stílusát.

Az FPDF elvileg nagyon hasonlít a HTML. Azok, akik már ismerik a HTML-t, könnyen létrehozhatnak bármilyen összetettségű PDF dokumentumot az FPDF segítségével, mivel a kód felépítése nagyon hasonló a HTML jelöléshez: a fejlécek, a szöveg, a képek és a táblázatok hozzáadása hasonló módon történik. Aki nem ismeri a HTML-t, annak nem kell aggódnia - használhatja az LLM, amely azonnal segít a kód összeállításában a kívánt dokumentum elrendezésének létrehozásához.

- A következő példa azt mutatja be, hogyan lehet egy jelentést létrehozni fejléccel és törzsszöveggel. Ha ezt a kódot bármelyik Python-támogatással rendelkező IDE-ben végrehajtjuk létrehoz egy PDF-fájlt, amely tartalmazza a kívánt fejlécet és szöveget:

```
from fpdf import FPDF      # Az FPDF könyvtár importálása
pdf = FPDF()    # PDF létrehozása -document
pdf.add_page()   # Oldal hozzáadása

pdf.set_font("Arial", style='B', size=16) # Betűtípus beállítása: Arial, félkövér, méret
                                             16
pdf.cell(200, 10, "Projektjelentés", ln=True, align='C') # Hozzon létre egy címet és
                                                         központosítsa azt.
pdf.set_font("Arial", size=12)  # A betűtípus módosítása normál Arial-ra, 12-es méretre
pdf.multi_cell(0, 10, "This document contains data on the results of project file verification...") #
Többsoros szöveg hozzáadása
pdf.output(r "C:\reports\report.pdf") # PDF mentése -file
```



7.2-14 ábra néhány sor Python kóddal automatikusan létrehozhatjuk a szükséges PDF szöveges dokumentumot.

A jelentések készítésekor fontos figyelembe venni, hogy a dokumentum alapjául szolgáló adatok ritkán statikusak. A fejlécek, szövegblokkok (7.2-14. ábra) gyakran dinamikusan alakulnak ki, az ETL-folyamatban a Transform szakaszában kapnak értékeket.

A kód használatával olyan dokumentumokat hozhat létre, amelyek naprakész információkat tartalmaznak: a projekt nevét, a jelentés létrehozásának dátumát, valamint a résztvevőkre vagy az aktuális állapotra vonatkozó információkat. A kódban lévő változók használata lehetővé teszi, hogy ezeket az adatokat automatikusan beilleszze a jelentés szükséges helyére, így teljesen kiküszöbölhető a kézi szerkesztés az elküldés előtt.

Az egyszerű szöveg és a címsorok mellett a táblázatok különleges helyet foglalnak el a projektdokumentációban. Szinte minden dokumentum tartalmaz strukturált adatokat: az

objektumleírásoktól a vizsgálati eredményekig. A táblázatok automatikus generálása a Transform szakasz adatai alapján nemcsak a dokumentumkészítés folyamatának felgyorsítását teszi lehetővé, hanem az információátvitel során felmerülő hibák minimalizálását is. Az FPDF lehetővé teszi a táblázatok beillesztését a PDF -fájlokba (szövegként vagy képként), a cellák határainak, oszlopméreteinek és betűtípusainak beállításával (7.2-15. ábra). Ez különösen akkor kényelmes, ha dinamikus adatokkal dolgozunk, amikor a sorok és oszlopok száma a dokumentumfeladatuktól függően változhat.

- A következő példa azt mutatja be, hogyan lehet automatizálni a táblázatok létrehozását, például az anyagok, becslések vagy paramétervizsgálatok eredményeinek listáját:

```
data = [
    ["Tétel", "Mennyiség", "Ár"], # oszlopcímek
    ["Beton", "10 m³", "$ 500."], # Első sor adatai
    ["Rebar", "2 tonna", "$ 600"], # Második sor adatai.
    ["Tégla", "5000 db", "$ 750."], # 3. sor adatai.
]

pdf = FPDF () # PDF létrehozása -document
pdf.add_page() # Oldal hozzáadása
pdf.set_font("Arial", size=12) # Betűtípus beállítása

for row in data: # Keresés a táblázat soraiban
    for item in row: # Menj végig a sor celláin
        pdf.cell(60, 10, item, border=1) # Létrehoz egy cellát 60-as szélességű és 10-es magasságú
        kerettel.
    pdf.ln() # Lépjünk a következő sorba
pdf.output(r "C:\reports\table.pdf") # PDF mentése -file
```

Item	Quantity	Price
Concrete	10 m³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

7.2-15 ábra A PDF PDF-ben a transzformációs lépésből nemcsak a szöveget, hanem bármilyen táblázatos információt is automatikusan létrehozhat a.

A valós jelentéstételi forgatókönyvekben a táblázatok általában dinamikusan generált információk, amelyeket az adattranszformációs szakaszban kapunk. Az adott példában (7.2-15. ábra) a táblázat statikus formában került be a PDF -dokumentumba: a példa adatai az adatszótárba kerültek (a kód első sora), valós körülmények között az ilyen adatváltozót automatikusan töltik ki pl. az adatkeret csoportosítása után.

A gyakorlatban az ilyen táblázatok gyakran a különböző dinamikus forrásokból származó strukturált adatok alapján épülnek fel: adatbázisok, Excel -fájlok, API -interfészek vagy analitikus számítások eredményei.

Leggyakrabban a Transform (ETL) szakaszban az adatokat aggregálják, csoportosítják vagy szűrik - és csak ezután alakítják át összegekké grafikonok vagy kétdimenziós táblázatok formájában, amelyek a jelentésekben megjelennek. Ez azt jelenti, hogy a táblázatok tartalma a kiválasztott paraméterektől, az elemzési időszaktól, a projekt szűrőitől vagy a felhasználói beállításoktól függően változhat.

A dinamikus adatkeretek és adatkészletek használata a Transform szakaszban a lehető legrugalmasabbá, skálázhatóvá és könnyen megismételhetővé teszi a jelentéstételi folyamatot a Load szakaszban, kézi beavatkozás nélkül.

A táblázatok és a szöveg mellett az FPDF támogatja a táblázatos adatok grafikonjainak hozzáadását is, ami lehetővé teszi a Matplotlib vagy más, fentebb tárgyalt vizualizációs könyvtárakkal generált képek beágyazását a jelentésbe. A kód segítségével bármilyen grafikon, diagram és diagram hozzáadható a dokumentumhoz.

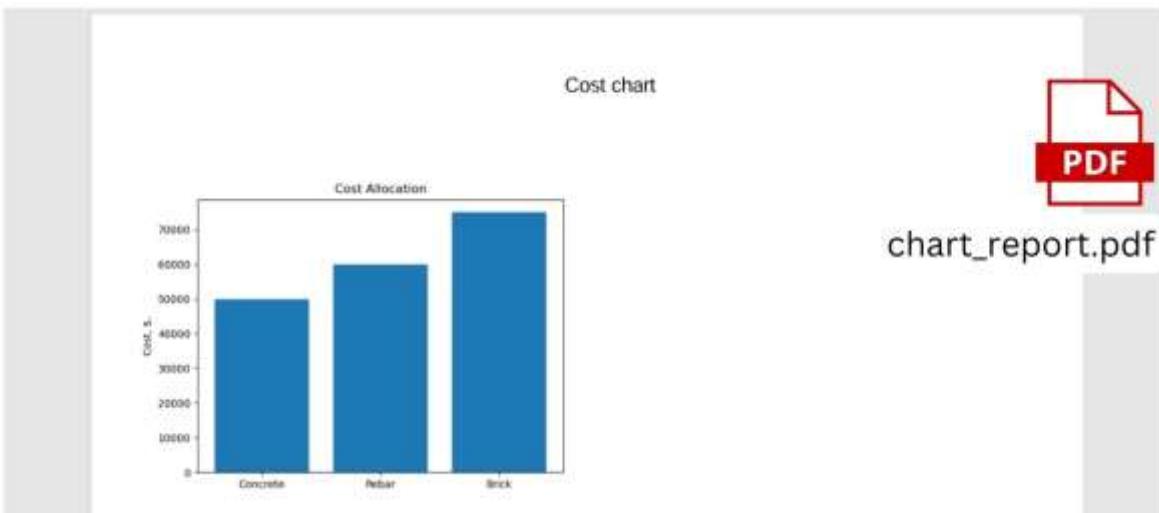
- A Python könyvtár FPDF segítségével adjunk hozzá egy Matplotlib. segítségével előzetesen generált grafikont a PDF dokumentumhoz:

```
import matplotlib.pyplot as plt # Import Matplotlib a grafikonok létrehozásához

fig, ax = plt.subplots() # Létrehozza a grafikon ábráját és tengelyeit.
categories = ["Beton", "Rebar", "Brick"] # Kategóriák nevei
values = [50000, 60000, 75000] # Kategóriák értékei
ax.bar(categories, values) # Sávdiagram létrehozása
plt.ylabel("Value,$.") # Jelöljük az Y tengelyt
plt.title("Költségeloszlás") # Cím hozzáadása
plt.savefig(r "C:\reports\chart\chart\chart\chart.png") # A diagram mentése képként

pdf = FPDF () # PDF létrehozása -dokumentum
pdf.add_page() # Oldal hozzáadása
pdf.set_font("Arial", size=12) # Betűtípus beállítása
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # Fejléc hozzáadása

pdf.image(r "C:\reports\chart\chart\chart\chart.png", x=10, y=30, w=100) # Kép
# beszúrása a PDF-be (x, y - koordináták, w - szélesség)
pdf.output(r "C:\reports\chart_report.pdf") # A PDF fájl mentése
```



7.2-16 Egy tucatnyi kódssorral létrehozhat egy grafikont, elmentheti, majd beillesztheti egy PDF-dokumentumba.

Az FPDF segítségével a dokumentumkészítés és a logika folyamata átláthatóvá, gyors és kényelmes lesz. A kódba épített sablonok lehetővé teszik a dokumentumok naprakész adatokkal történő generálását, kiküszöbölvé a kézi kitöltés szükségességét.

ETL-automatizálás használata - az időigényes kézi jelentéskészítés helyett a szakemberek az adatok elemzésére és a döntéshozatalra összpontosítanak, ahelyett, hogy a megfelelő eszközök választanák ki egy adott adatsíróval való munkához, egyértelmű felhasználói felülettel.

Így az FPDF könyvtár rugalmas eszközök biztosít bármilyen összetettségű dokumentum automatikus létrehozásához - a rövid műszaki jelentésektől a táblázatokkal és diagramokkal ellátott összetett analitikus összefoglalókig, ami nemcsak a dokumentumáramlás felgyorsítását teszi lehetővé, hanem jelentősen csökkenti a kézi adatbevitellel és formázással kapcsolatos hibák valószínűségét is.

ETL Load: Jelentések készítése és betöltése más rendszerekbe.

A betöltési szakaszban az eredményeket táblázatok, grafikonok és a meghatározott követelményeknek megfelelően elkészített PDF formátumú végleges jelentések formájában állították elő. Lehetőség van továbbá ezen adatok gépileg olvasható formátumba (pl. CSV) történő exportálására, ami szükséges a külső rendszerekkel, például ERP, CAFM, CPM, BI platformokkal és egyéb vállalati vagy ipari megoldásokkal való integrációhoz. A CSV mellett feltöltés történhet XLSX, JSON, XML formátumba vagy közvetlenül az automatikus információtároló adatbázisokba is.

- ❸ A betöltési lépés automatizálásához szükséges megfelelő kód generálásához egyszerűen kérdezze le például az LLM -interfészét: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude vagy QWEN:

Írjon kódot az adatellenőrzés eredményeinek DataFrame -ban történő jelentés létrehozására, ahol az "verified_" előtaggal ellátott oszlopokat megszámolja, átnevezi "Passed" és "Failed" (megfelelt) és "Failed" (nem felelt meg), a hiányzó értékeket 0-val helyettesíti, majd csak azokat a sorokat exportálja egy CSV -fájlba, amelyek minden érvényesítésen megfeleltek.

2 LLM válasza:



The screenshot shows a code editor window titled "Pipeline.py". The code is as follows:

```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

At the top right of the editor, there is a "Create in LLM Chat" button with four icons: a brain, a document, a star, and a whale. Below the editor, there are four small icons: a terminal, a cloud, a PC, and a file. A "Run in IDE" button is located at the bottom right of the editor area.

7.2-17. ábra A transzformációs lépésekben a végleges adatkeretből nyert hitelesített adatokat más rendszerekkel való integrálás céljából egy CSV -fájlba exportáljuk.

A fenti kód (7.2-17. ábra) az ETL -folyamat utolsó szakaszát - a betöltést - valósítja meg, amelynek során az ellenőrzött adatok CSV formátumban kerülnek elmentésre, amely kompatibilis a legtöbb külső rendszерrel és adatbázissal. Így az ETL -folyamat teljes ciklusát elvégeztük, beleértve az adatok kinyerését, átalakítását, megjelenítését, dokumentálását és exportálását a szükséges rendszerekbe és formátumokba, ami biztosítja az információkkal való munka reprodukálhatóságát, átláthatóságát és automatizálását.

Az ETL csővezeték egyaránt használható egyedi projektek és nagyszabású alkalmazások feldolgozására - amikor dokumentumok, képek, szkennelések, CAD projektek, pontfelhők, PDF fájlok vagy más, elosztott rendszerekből származó források formájában beérkező adatok százait vagy ezreit elemzi. A folyamat teljes automatizálásának képessége az ETL-t nem csupán technikai feldolgozóeszközé, hanem a digitális építőipari információs infrastruktúra alapjává teszi.

ETL az LLM-mel: adatok vizualizálása PDF -dokumentumokból

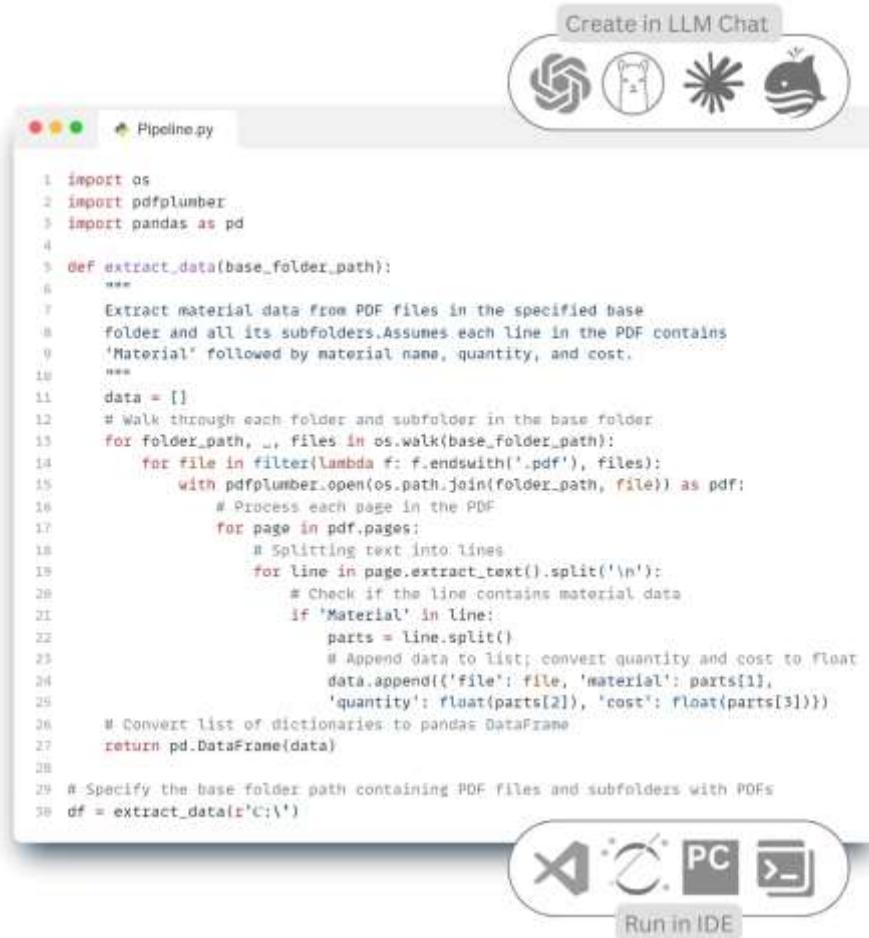
Itt az ideje, hogy áttérjünk egy teljes körű ETL-folyamat kialakítására, amely az adatkezelés minden kulcsfontosságú szakaszát egyetlen forgatókönyvben - az extrakciót, az átalakítást és a betöltést - lefedi. Építsünk egy automatizált ETL-csatornát, amely lehetővé teszi a PDF-dokumentumok kézi munka nélküli feldolgozását - az adatok kivonása a dokumentumokból, vizualizálása, elemzése és más rendszerekbe történő átvitele.

Példánkban az ETL-folyamatot felszólításokon keresztül írjuk le, amelyeknek a nyelvi modell (LLM) számára az összes ETL-folyamatot el kell magyarázniuk a végeredmény leírásával együtt. Ebben az esetben a feladat a megadott mappában és almappában található összes PDF-fájl megkeresése, a releváns információk - pl. anyagnevek, mennyiségek és költségek - kinyerése belőlük, és az eredmény strukturált táblázat (DataFrame) formájában történő bemutatása a további elemzéshez

- ❶ Első szöveges lekérdezés az LLM oldalon, hogy automatikusan kivonatolja az adatokat több PDF dokumentumból és létrehozza az adatok adatkeretét az Extract lépéshoz:

Írjon kódot az anyaginformációk kinyerésére egy adott mappában és annak almappáiban található PDF fájlokból. A PDF-ben található adatok között szerepel az anyag neve, mennyisége és költsége. Az eredményt egy DataFrame-ba kell menteni ↵

- ❷ LLM válasza :



```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26    # Convert list of dictionaries to pandas DataFrame
27    return pd.DataFrame(data)
28
29 # Specify the base folder path containing PDF files and subfolders with PDFs
30 df = extract_data(r'C:\')

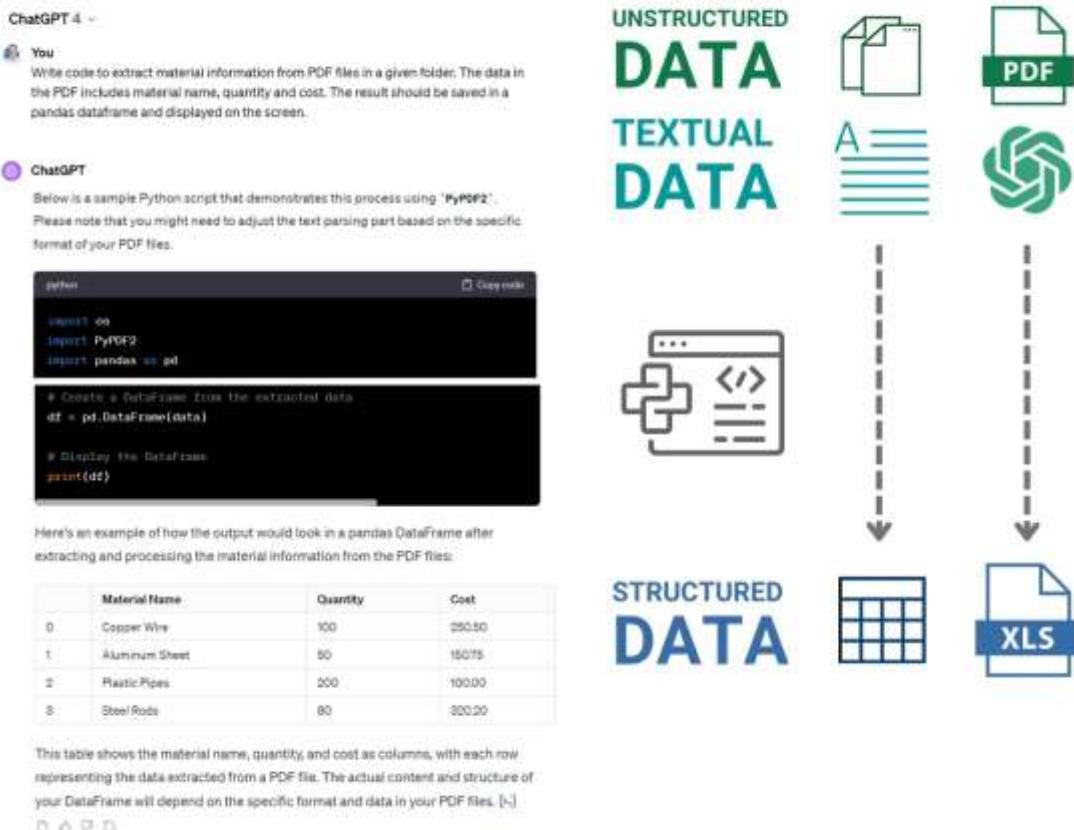
```

7.2-18 ábra LLM Python -kódot hoz létre egy adott mappában és annak összes almappájában található PDF -fájlból történő adatkiragadáshoz.

Az LLM válasz (7.2-18. ábra) egy kész Python -script, amely automatikusan végigjárja az összes mappát,

megnyitja a talált PDF -fájlokat, szöveges információkat nyer belőlük, és táblázatba konvertálja őket. Az így kapott csevegő kód futtatható a népszerű IDE-k valamelyikében: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse a PyDev pluginnal, Thonny, Wing IDE, IntelliJ IDEA a Python pluginnal, JupyterLab vagy a népszerű online eszközök: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternatívaként, a gyűjtési folyamat egyszerűsítése érdekében, ahelyett, hogy a kódot az LLM oldalról másolnánk ki és az IDE-ben használnánk, egy tucat PDF -fájlt közvetlenül az LLM chatre is feltölthetünk (7.2-19. ábra), és kimenetként egy táblázatot kapunk, anélkül, hogy a kódot látnánk vagy futtatnánk. A kód végrehajtásának eredménye egy táblázat lesz az általunk kiválasztott attribútumokkal.

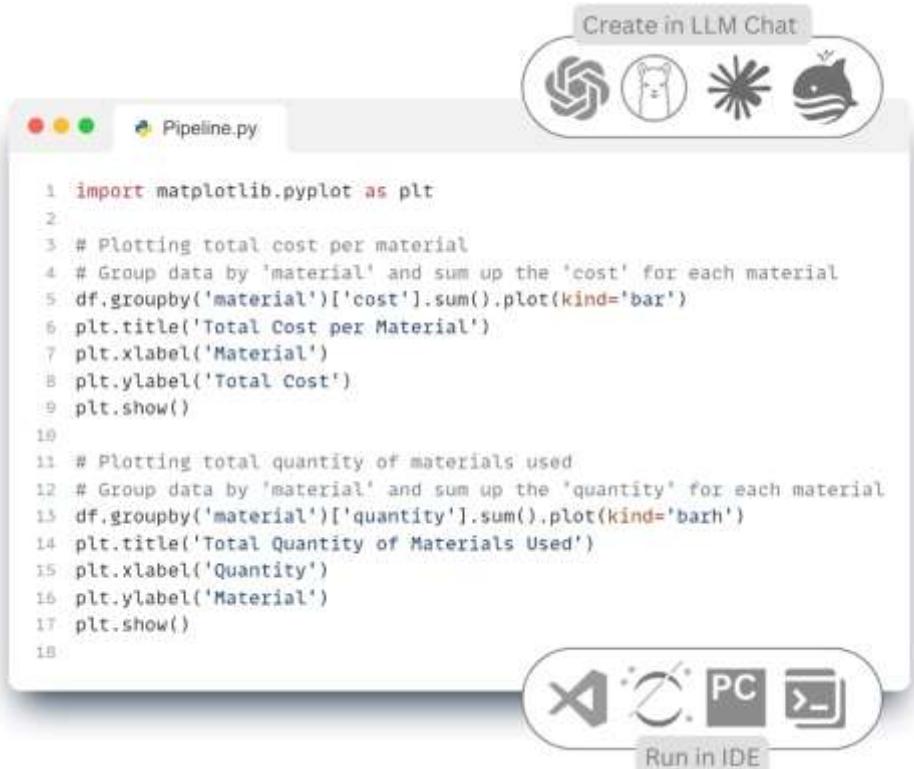


7.2-19. ábra Az LLM -ben található kód végrehajtásának eredménye, amely a PDF -fájlok adatait egy strukturált adatkeret nézetben, kiválasztott attribútumokkal kivonatolja.

A következő lépésben nyelvi modellt kérünk a kapott adatokra - például az anyagfelhasználás költségeinek és mennyiségének összehasonlítására, és néhány olyan példavizualizáció létrehozására, amely a további elemzés alapjául szolgál.

- ⌚ Kérdezze meg az LLM folyamatos beszélgetésben, hogy a transzformációs lépésben készített táblázatokból (7.2-18. ábra) rajzoljon fel néhány grafikont a transzformációs lépésben készített táblázatokból (7.2-18. ábra):

Az egyes anyagok összköltségének és mennyiségeinek megjelenítése a adatmezőből (7.2-18. ábra)



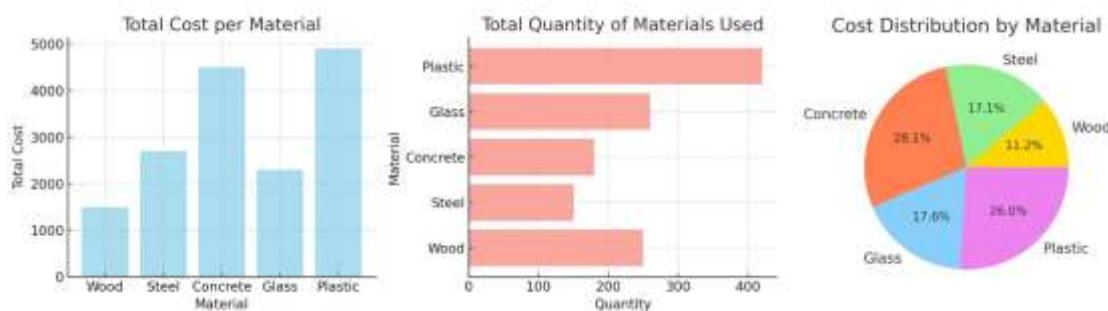
```

1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

Ábra 7.2-20 LLM válasz -modellek Python kódként adatok megjelenítéséhez egy adatkeretből a matplotlib könyvtár segítségével.

Az LLM automatikusan generálja és végrehajtja a Python -kódot (7.2-20. ábra) a matplotlib könyvtár segítségével. E kód végrehajtása után közvetlenül a chatben kapjuk meg az építési projektek költségeinek és anyagfelhasználásának grafikonjait (7.2-21. ábra), ami nagyban leegyszerűsíti az elemző munkát.

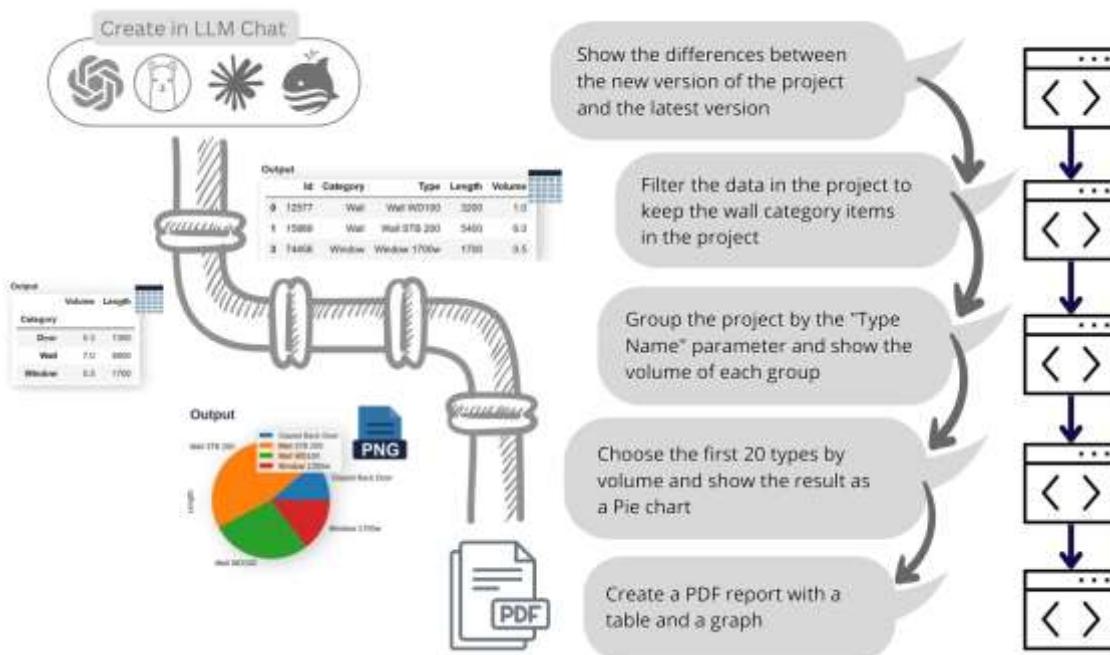


Ábra 7.2-21 Az LLM-válasz vizualizációja grafikonok formájában a DataFrame-ben gyűjtött adatok alapján.

Az ETL kód megírásához, a kód elemzéséhez és végrehajtásához, valamint az eredmények vizualizálásához

szükséges ötletek kidolgozásához az LLM egyszerű szöveges lekérdezéseken keresztül kapható támogatás, anélkül, hogy a programozás alapjait meg kellene tanulni. A mesterséges intelligencia eszközök, például az LLM megjelenése határozottan megváltoztatja a programozás és az adatfeldolgozás automatizálásának megközelítését (7.2-22. ábra).

A PwC jelentése szerint "Mi a mesterséges intelligencia valódi értéke a vállalkozásod számára, és hogyan tudod kihasználni?". (2017) [139] szerint a folyamatok automatizálása és a termelékenység javítása lesz a gazdasági növekedés fő hajtóereje. A termelékenység javulása pedig várhatóan az összes mesterséges intelligencia által vezérelt GDP-növekedés több mint 55%-át teszi majd ki 2017 és 2030 között".



7.2-22 AI LLM ábra Segít olyan kódtervezetet létrehozni, amelyet a későbbi projektekben LLM nélkül alkalmaznak.

Az olyan eszközökkel, mint a ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok, valamint a nyílt adatok és a nyílt forráskódú szoftverek, olyan folyamatokat automatizálhatunk, amelyeket korábban csak speciális, magas költségű és nehezen karbantartható, moduláris, saját fejlesztésű rendszerekkel végezhettünk.

Az építőiparban ez azt jelenti, hogy azok a vállalatok, amelyek elsőként vezetik be az automatizált Pipeline -adatfolyamatokat, jelentős előnyökre tesznek szert, a projektmenedzsment hatékonyságának növelésétől kezdve a pénzügyi veszteségek csökkentésén át a széttagolt alkalmazások és a silózott adattárházak megszüntetéséig.

Az ETL-folyamatban az üzleti feladatok végrehajtásának leírt logikája az analitika és az adatfeldolgozási folyamatok automatizálásának alapvető része, amely egy tágabb fogalom - a csővezetékek (Pipelines) - sajátos változata.

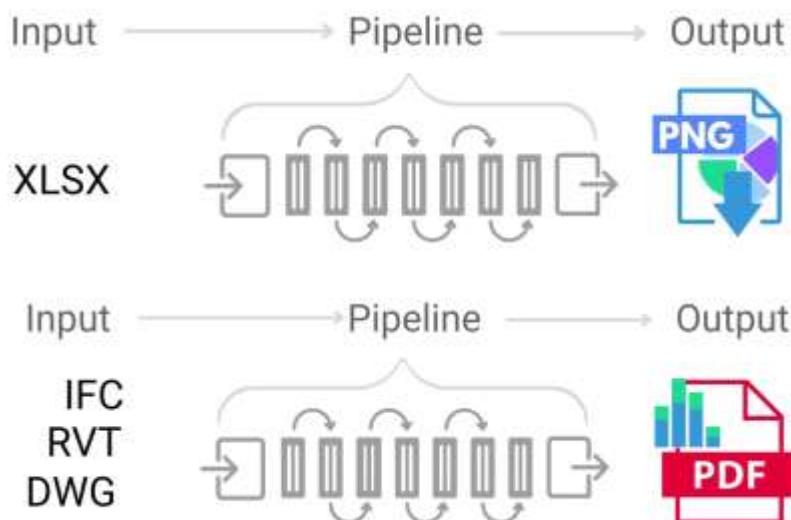
FEJEZET 7.3.

AUTOMATIKUS ETL SZÁLLÍTÓSZALAG (CSŐVEZETÉK)

Pipeline: Automatikus ETL szállítószalag adatok

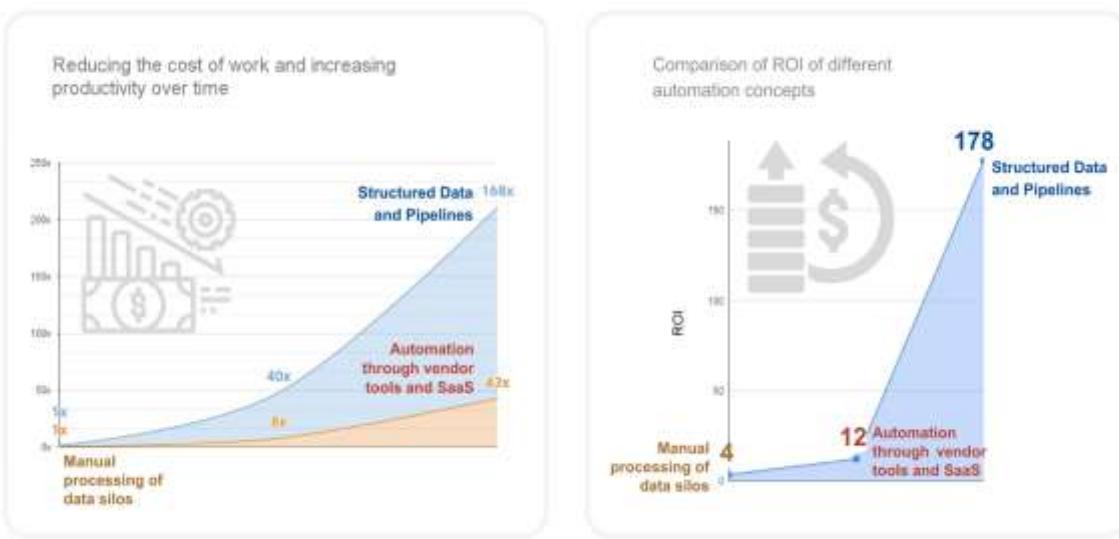
Az ETL-folyamatot hagyományosan az analitikai rendszerek adatfeldolgozására használják, mind a strukturált, mind a strukturálatlan forrásokra kiterjedően. A mai digitális környezetben azonban egyre gyakrabban használnak egy tágabb kifejezést - Pipeline (futószalag), amely a feldolgozás bármely olyan szekvenciális láncát írja le, ahol az egyik szakasz kimenete a következő szakasz bemenetévé válik.

Ez a megközelítés nemcsak az adatokra, hanem az automatizálás más típusaira is vonatkozik: feladatfeldolgozás, jelentéskészítés, szoftverekkel való integráció és digitális munkafolyamatok (7.3-1. ábra).



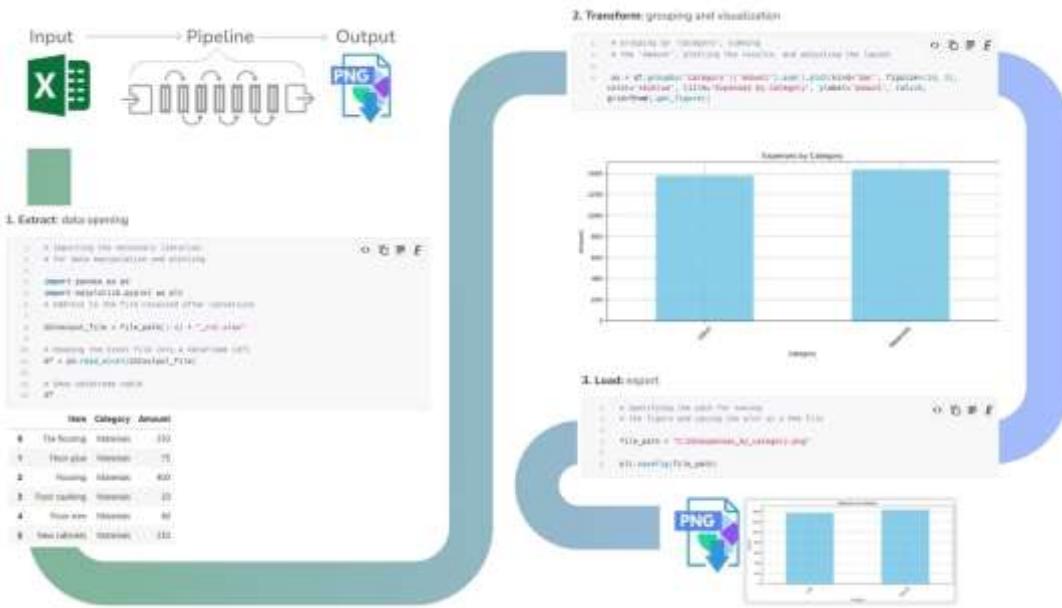
7.3-1. ábra Pipeline egy olyan feldolgozási sorrend, amelyben az egyik szakasz kimenete a következő szakasz bemenetévé válik.

csővezeték használata az automatizálás egyik fő eleme, különösen, ha nagy mennyiségű heterogén adattal dolgozunk. A pipeline-architektúra lehetővé teszi az összetett feldolgozási lépések moduláris, konzisztens és kezelhető formátumba szervezését, ami növeli az olvashatóságot, egyszerűsíti a kód karbantartását, és lehetővé teszi az inkrementális hibakeresést és a skálázható tesztelést.



Ábra 7.3-2 ROI Pipeline adatérvényesítési folyamat több tízszeresére és százszorosára csökkenti a végrehajtási időt a klasszikus eszközökkel történő feldolgozáshoz képest [74].

A saját rendszerekben (ERP, PMIS, CAD, stb.) végzett kézi munkával szemben a pipelining lehetővé teszi a feladatok sebességének jelentős növelését (7.3-2. ábra), az ismétlődő munka elkerülését és a folyamatok megfelelő időben történő elindításának automatizálását (7.3-3. ábra).

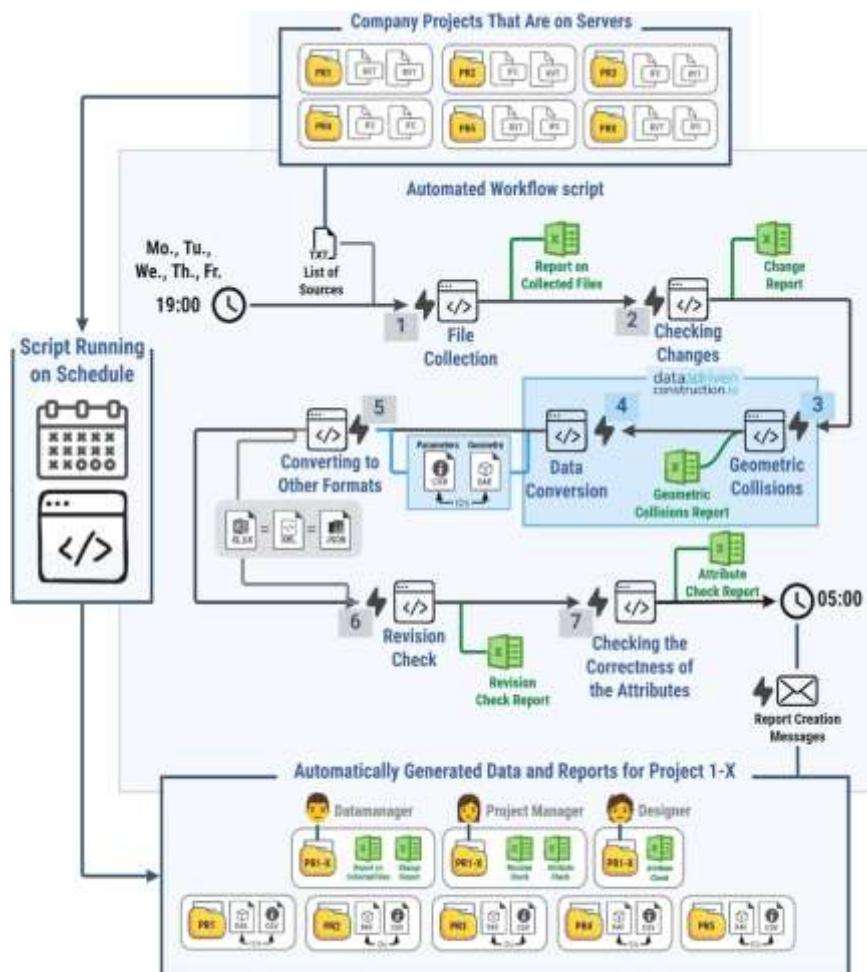


Ábra 7.3-3 ETL példa csővezeték egy XLSX fájlban lévő táblázatos adatokból automatikusan grafikonok készítésére az Excel megnyitása nélkül.

Az áramló adatok feldolgozásához és az ETL-folyamathoz hasonló automatizált csővezeték létrehozásához előre meg kell határoznia az adatforrásokat, valamint az adatgyűjtés időkeretét - akár egy adott üzleti folyamatra, akár az egész vállalatra vonatkozóan.

Az építési projektekben az adatok számos heterogén forrásból származnak, különböző frissítési időközökkel. A megbízható adatbemutató létrehozása érdekében kritikus fontosságú az információk lekérdezésének és frissítésének időpontját rögzíteni. Ez lehetővé teszi az időben történő döntéshozatalt és javítja a projektmenedzsment hatékonyságát.

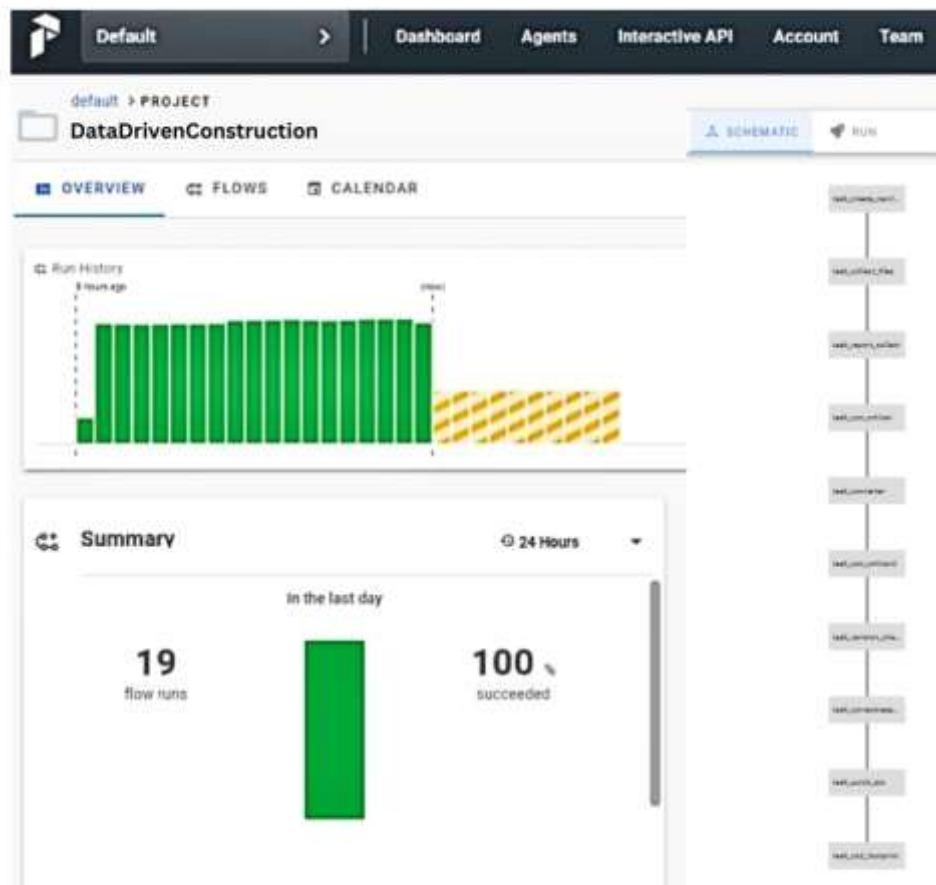
Az egyik lehetőség, hogy az összeszerelési folyamatot egy meghatározott időpontban, például 19:00-kor, a munkanap végén - indítják el. Ekkor aktiválódik a különböző rendszerekből és tárolókból származó adatok összesítéséért felelős első szkript (7.3-4. ábra 1. lépés). Ezt követi az adatok automatikus feldolgozása és átalakítása az elemzésre alkalmas strukturált formátumba (7.3-4. ábra 2-4. lépés). A végső szakaszban az előkészített adatok felhasználásával automatikusan generálódnak a jelentések, műszerfalak és az előző fejezetekben ismertetett egyéb termékek (7.3-4. ábra 6-7. lépés). Ennek eredményeként a vezetők reggel 05:00 órakor már a kívánt formátumban naprakész projektállapot-jelentésekkel rendelkeznek (7.3-5. ábra).



Ábra 7.3-4 A csővezetékben lévő, este automatikusan összegyűjtött adatokat éjszaka feldolgozzák, hogy a vezetők reggelre naprakész jelentésekkel és friss jelentésekkel rendelkezzenek.

Az időben történő adatgyűjtés, a KPI-k meghatározása, az átalakítási folyamatok automatizálása és a műszerfalakon keresztül történő megjelenítés a sikeres adatvezérelt döntéshozatal kulcsfontosságú elemei.

Az ilyen automatizált folyamatok (7.3-4. ábra) teljes autonómiával végrehajthatók: ütemezés szerint futnak, felügyelet nélkül dolgozzák fel az adatokat, és akár a felhőben, akár a vállalat saját szerverén telepíthetők (7.3-5. ábra). Ez lehetővé teszi, hogy az ilyen ETL-csővezetékek integrálhatók legyenek a meglévő IT-infrastruktúrába, megőrizve az adatok feletti ellenőrzést és rugalmasan skálázhatók.



Ábra: 7.3-5 Automatikus ETL -szállító folyamatok (7.3-4. ábra) a Prefect platformon, amelyben 10 python szkriptet futtatnak felváltva minden munkanapon 19:00 után.

A munkafolyamatok automatizálása nemcsak a csapatok termelékenységét növeli azáltal, hogy időt szabadít fel az értelmesebb és kevésbé rutinszerű feladatokra, hanem fontos első lépésként szolgál a mesterséges intelligencia (AI) technológiák üzleti folyamatokba való beépítése felé, amelyet a prediktív analitikáról és a gépi tanulásról szóló fejezetben tárgyalunk részletesebben.

Csővezeték -ETL adatérényesítési folyamat LLM-mel

Az előző, az adatkövetelmények létrehozásáról és az ETL automatizálásáról szóló fejezetekben lépésről lépésre lebontottuk az adatelőkészítés, -átalakítás, -érvényesítés és -megjelenítés folyamatát. Ezeket a tevékenységeket különálló kódblokkokkókként valósítottuk meg (7.2-18. ábra - 7.2-20. ábra), amelyek mindegyike egy adott feladatot hajt végre.

Most következik a következő cél: ezeket az elemeket egyetlen, összefüggő és automatizált adatfeldolgozási csővezetékbe - egy csővezetékbe, ETL -Pipeline -egyesíteni, amelyben az összes szakasz (betöltés, validálás, vizualizáció, exportálás) egyetlen automatikusan végrehajtható szkriptben egymás után hajtódiák végre.

A következő példában egy teljes adatfeldolgozási ciklus kerül végrehajtásra: a forrás CSV fájl betöltésétől →a struktúra és az értékek ellenőrzéséig a reguláris kifejezések segítségével →az eredmények kiszámításáig→ egy vizuális jelentés generálása PDF formátumban.

- ❷ Az alábbi szöveges lekérdezéssel az LLM. segítségével lekérdezheti a megfelelő kódot:

Kérjük, írjon egy olyan kódmintát, amely betölti az adatokat a CSV-ból, ellenőrzi a DataFrame-et a reguláris kifejezések segítségével, ellenőrzi az azonosítókat 'W-NEW' vagy 'W-OLD' formátumban, az energiahatékonyságot 'A' és 'G' betűkkel, a garanciális időszakot és a csereciklust években megadott számértékekkel, és a végén létrehoz egy jelentést a megfelelő és nem megfelelő értékek számával, létrehoz egy PDF-et az eredmények hisztogramjával és egy szöveges leírással. ↵

■ LLM válasza :

```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$',
14     'Energy_Efficiency': r'^[A-G]\+*$',
15     'Warranty_Period': r'^\d+\syears$',
16     'Replacement_Cycle': r'^\d+\syears$'
17 }
18
19 for column, pattern in patterns.items():
20     df[f'verified_{column}'] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38 # Add text page
39 plt.figure()
40 plt.axis('off')
41 text_str = "Verification Results Summary"
42 plt.text(0.1, 0.5, text_str, wrap=True)
43 pdf.savefig()
44 plt.close()
45
46 pdf_path
47

```

Ábra: 7.3-6 Pipeline (ETL) automatizálja az adatfeldolgozás teljes ciklusát: a letöltéstől és az érvényesítéstől a strukturált jelentés PDF formátumban történő létrehozásáig .

Az automatizált kód (7.3-6. ábra) az LLM csevegőszobában vagy a DIE-ben a kód másolása után a megadott reguláris kifejezésekkel -fájlból származó adatokat a CSV -fájlból validálja, jelentést készít a sikeres és sikertelen rekordok számáról, majd az érvényesítési eredményeket PDF -fájlként elmenti.

Ez az ETL -szállítói struktúra, ahol minden egyes lépés - az adatbetöltéstől a jelentéskészítésig - külön modulként van megvalósítva, biztosítja az átláthatóságot, a skálázhatóságot és a reprodukálhatóságot. Az érvényesítési logika könnyen olvasható Python-kódként való bemutatása nemcsak a fejlesztők, hanem az adatkezelési, minőségügyi és elemzési szakemberek számára is átláthatóvá és érhetővé teszi a folyamatot.

A Pipeline -nak az adatfeldolgozás automatizálására vonatkozó megközelítése lehetővé teszi a folyamatok szabványosítását, ismételhetőségük növelését és az új projektekre való alkalmazkodás egyszerűsítését. Ez egységes módszertant hoz létre az adatok elemzésére, függetlenül a forrástól vagy a feladat típusától - legyen szó megfelelési tesztelésről, jelentéstételről vagy külső rendszerekbe történő adatátvitelről.

Az ilyen automatizálás csökkenti az emberi hibák számát, csökkenti a saját megoldásoktól való függőséget, valamint növeli az eredmények pontosságát és megbízhatóságát, ami alkalmassá teszi őket minden a projekt szintű operatív analitikára, minden a vállalati szintű stratégiai analitikára.

Pipeline -ETL: a projektelemek adatainak és információinak ellenőrzése a CAD-ben (BIM).

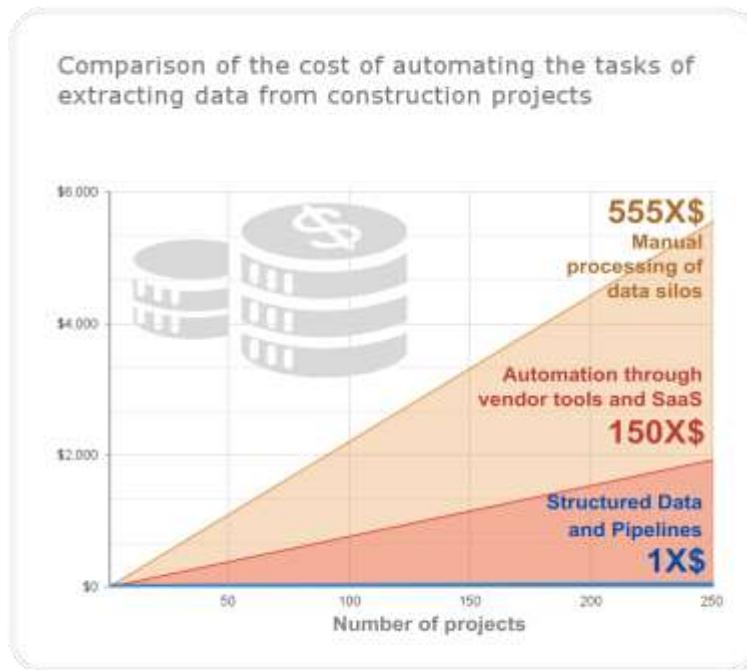
A CAD-rendszerekből és adatbázisokból származó adatok (BIM) az építőipari vállalatok üzleti életében a legkifinomultabb és legdinamikusabban frissülő adatforrások közé tartoznak. Ezek az alkalmazások nemcsak geometriai adatokkal írják le a projektet, hanem többrétegű szöveges információkkal is kiegészítik azt: térfogatok, anyagtulajdonságok, helyiségekbeosztások, energiahatékonysági szintek, túrések, életciklusok és egyéb attribútumok.

A CAD -modellekben az entitásokhoz rendelt attribútumok a tervezési szakaszban alakulnak ki, és a további üzleti folyamatok alapjává válnak, beleértve a költségszámítást, az ütemezést, az életciklus-értékelést és az ERP és CAFM -rendszerekkel való integrációt, ahol a folyamatok hatékonysága nagyban függ a tervezési részlegektől származó adatok minőségétől.

A CAD- (BIM-) modellek attribútumainak validálásának hagyományos megközelítése a kézi validálást foglalja magában (7.2-1. ábra), ami hosszú és költséges folyamattá válik, ha a modellek mennyisége nagy. Tekintettel a modern építési projektek mennyiségrére és számára, valamint rendszeres frissítéseikre, az adatok validálásának és átalakításának folyamata fenntarthatatlanná és megfizethetetlenné válik.

A főtállalkozóknak és a projektmenedzsereknek nagy mennyiségű projektadatot kell feldolgozniuk, beleértve ugyanazon modellek több változatát és töredékét. Az adatok RVT, DWG, DGN, IFC, NWD és egyéb formátumokban érkeznek a tervező szervezetektől (3.1-14. ábra), és rendszeresen ellenőrizni kell, hogy megfelelnek-e az ipari és vállalati szabványoknak

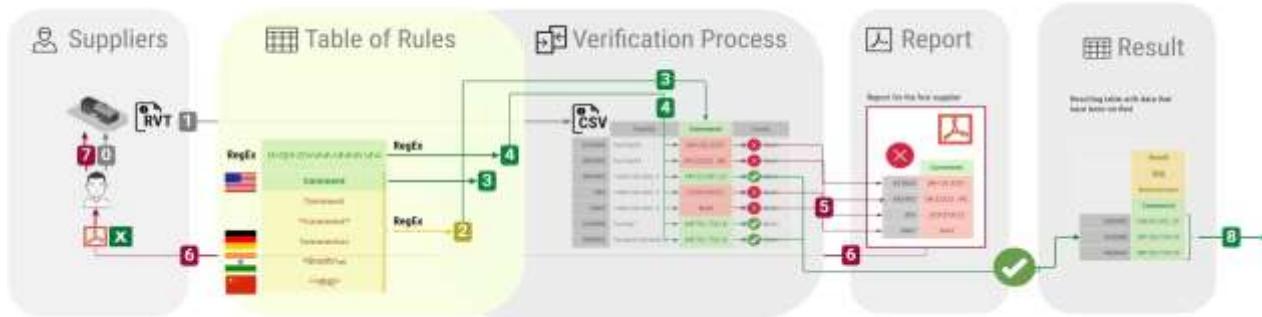
A kézi műveletektől és a speciális szoftverektől való függés miatt az adatérényesítési folyamat szűk keresztmetszetet jelent a vállalati szintű modellekből származó adatokkal kapcsolatos munkafolyamatokban. Az automatizálás és a strukturált követelmények használata kiküszöbölte ezt a függőséget, és ezáltal jelentősen megnövelte az adatérényesítés sebességét és megbízhatóságát (7.3-7. ábra).



Ábra 7.3-7 Az automatizálás növeli az adatok ellenőrzésének és feldolgozásának sebességét, ami több tucatszorosára csökkenti a munka költségét [140].

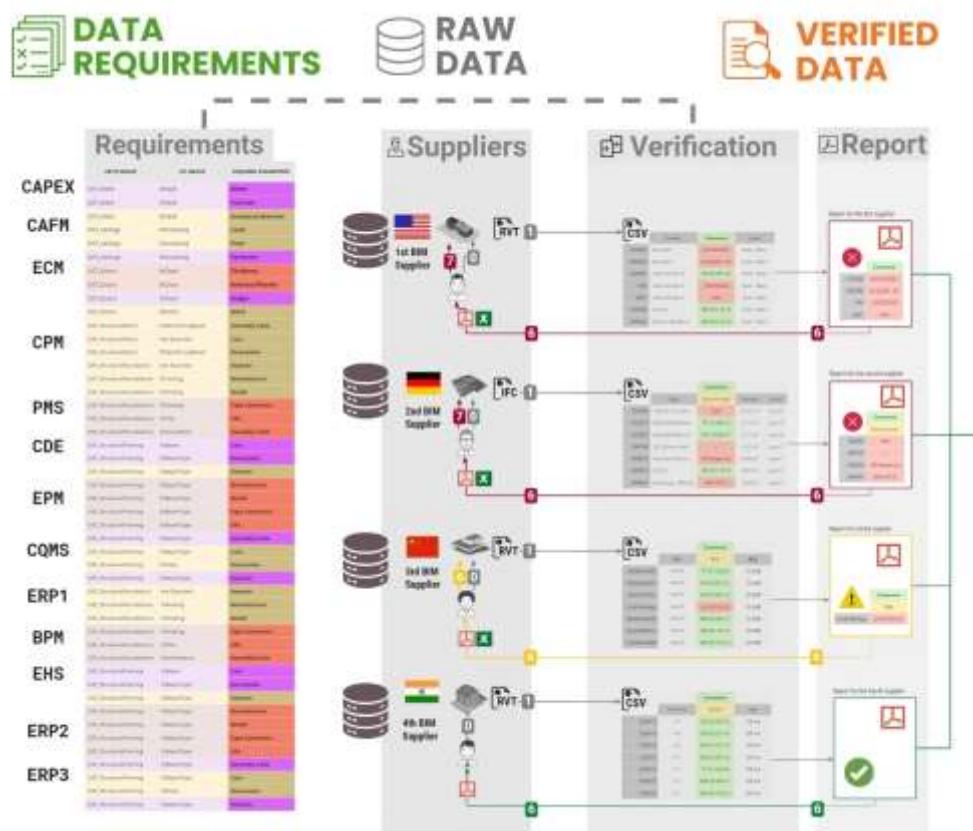
A CAD-adatok érvényesítési folyamata magában foglalja az adatok kinyerését (ETL szakasz Extract) különböző zárt (RVT, DWG, DGN, NWS stb.) vagy nyílt félig strukturált és parametrikus formátumokból (IFC, CPXML, USD.) vagy nyílt félig strukturált és parametrikus formátumokból (IFC, CPXML, USD), amelyekben az egyes attribútumokra és azok értékeire (Transform szakasz) szabályos kifejezésekkel RegEx (7.3-8. ábra) szabálytáblákat lehet alkalmazni, amely folyamatot a könyv negyedik részében részletesen tárgyaltuk.

és a sikeresen hitelesített rekordok PDF hibajelentésének létrehozását olyan strukturált formátumú kimenettel (betöltési lépés) kell befejezni, amely csak a további folyamatokhoz felhasználható hitelesített entitásokat veszi figyelembe.



7.3-8. ábra A projekt adatszolgáltatóitól a szabályos kifejezésekkel hitelesített végső jelentésig tartó adatérvényesítési folyamat.

A CAD-rendszerekből származó adatok validálásának automatizálása (BIM) strukturált követelményekkel és az ETL-csővezetékeken (7.3-9. ábra) keresztül feldolgozott új adatok áramoltatásával csökkenti a validálási folyamatba való kézi beavatkozás szükségességét (az egyes validálási és adatkövetelményfolyamatokat az előző fejezetekben tárgyaltuk).



Ábra 7.3-9 Az adatérvényesítés automatizálása az ETL segítségével egyszerűsíti az építési projektek irányítását a folyamatok felgyorsításával.

Hagyományosan a vállalkozók és a CAD (BIM) szakemberek által biztosított modellek validálása napokig

vagy hetekig is eltarthat. Az automatizált ETL-folyamatok bevezetésével azonban ez néhány percre csökkenthető. Egy tipikus helyzetben a kivitelező megállapítja: "A modell validálva van és megfelel a követelményeknek". Ez a kijelentés elindítja a vállalkozó adatminőségi állításának ellenőrzési láncolatát:

- ⌚ Projektmenedzser - "A vállalkozó azt állítja: "A modellt teszteltük, minden rendben van"."
- ⌚ Adatkezelő - Betöltésérvényesítés:
 - ▀ Egy egyszerű Pandas szkript másodpercek alatt észleli a jogosértést. Az automatizálás kiküszöböli a vitákat:
 - Kategória: OST_StructuralColumns, Paraméter: FireRating IS NULL.
 - A jogosértések azonosítóinak listájának létrehozása → exportálása Excel/PDF formátumba.

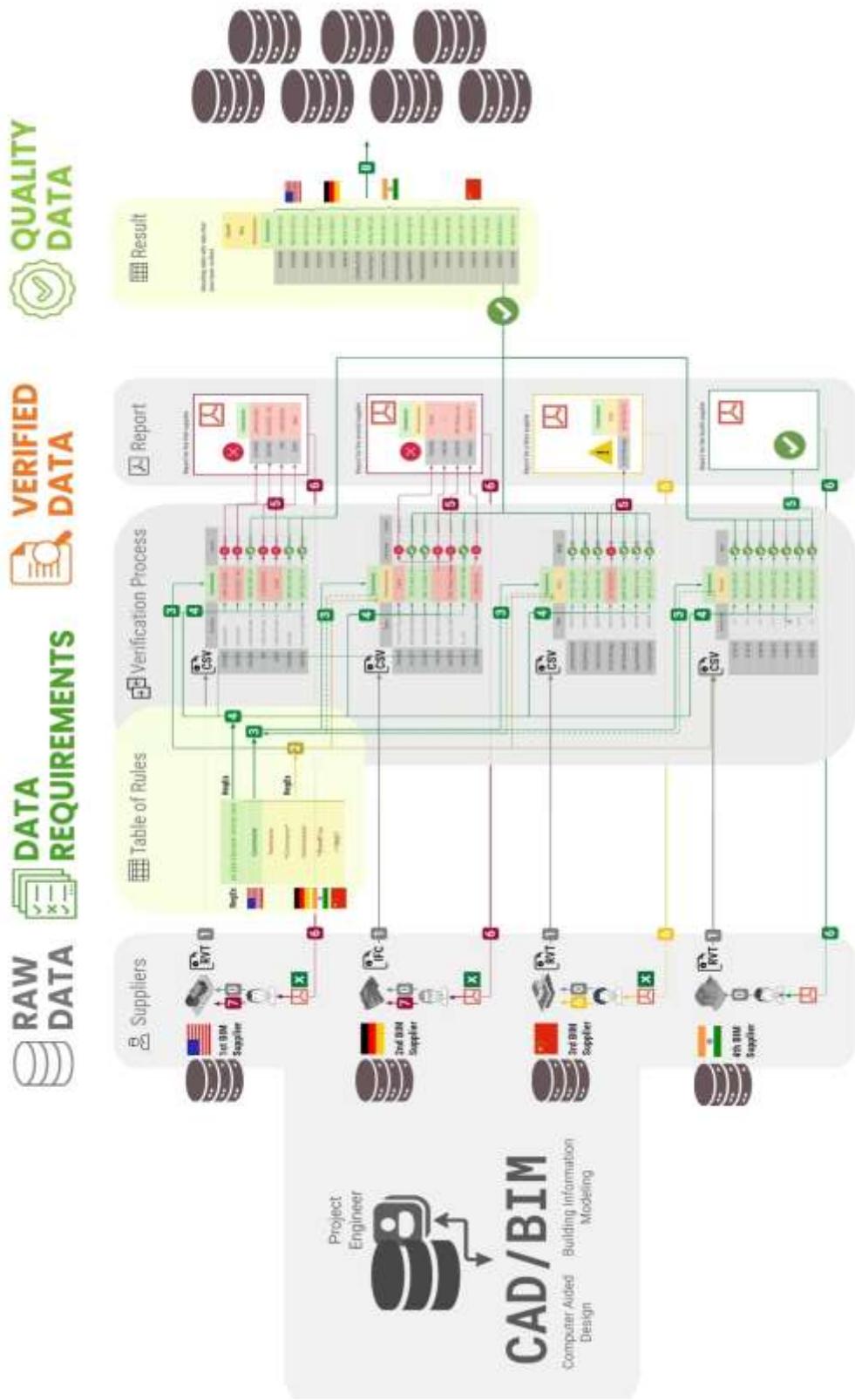
Egy egyszerű Pandas szkript másodpercek alatt észleli a jogosértést:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Szűrés
issues = df[df["FireRating"].isnull()] # Üres értékek
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # ID-k exportálása
```

- ⌚ Adatkezelő a projektmenedzsernek - "A oldal ellenőrzése azt mutatja, hogy 18 oszlopban nincs kitöltve a FireRating paraméter."
- ⌚ A projektvezető a vállalkozónak - "A modellt felülvizsgálatra visszaküldjük: a FireRating paraméter kötelező, nélküle az átvétel nem lehetséges."

Ennek eredményeként a CAD-modell nem esik át validáláson, az automatizálás kiküszöböli a vitákat, és a vállalkozó szinte azonnal kap egy strukturált jelentést a problémás elemek azonosítóinak listájával. Ily módon a validálási folyamat átláthatóvá, megismételhetővé és az emberi hibáktól védetté válik (7.3-10. ábra).

Ez a megközelítés az adatok érvényesítési folyamatát kézi minőségellenőrzés helyett mérnöki funkcióvá alakítja. Ez nemcsak a termelékenységet növeli, hanem lehetővé teszi azt is, hogy ugyanazt a logikát alkalmazzák a vállalat összes projektjére, lehetővé téve a folyamatok végponttól végpontig tartó digitális átalakítását a tervezéstől az üzemeltetésig.



7.3-10 Az elemattribútumok ellenőrzésének automatizálása kiküszöböli az emberi hibát és csökkenti a hibák valószínűségét.

Az automatizált csővezetékek (7.3-10. ábra) segítségével a CAD- (BIM-) rendszerektől minőségi adatokat váró rendszerhasználók azonnal megkaphatják a szükséges kimeneti adatokat - táblázatok, dokumentumok, képek - és gyorsan integrálhatják azokat a munkafeladataikba.

Az ellenőrzés, a feldolgozás és az elemzés automatizálása változást hoz az építési projektmenedzsment megközelítésében, különösen a különböző rendszerek átjárhatosága terén, anélkül, hogy bonyolult és drága moduláris, saját fejlesztésű rendszereket vagy zárt gyártói megoldásokat kellene használni.

Miközben a fogalmak és a marketinges rövidítések jönnek és mennek, maguk az adatkövetelmények érvényesítési folyamatai örökre az üzleti folyamatok szerves részei maradnak. Ahelyett, hogy újabb és újabb speciális formátumokat és szabványokat hoznának létre, az építőiparnak inkább olyan eszközökre kellene támaszkodnia, amelyek más iparágakban már bizonyították hatékonyságukat. Ma már léteznek olyan hatékony platformok az adatfeldolgozás és a folyamatintegráció automatizálására, amelyek lehetővé teszik a vállalatok számára a rutinműveletekre fordított idő jelentős csökkentését és a hibák minimalizálását az Extract, Transform és Load oldalakon.

Az ETL folyamatok automatizálására és összehangolására szolgáló megoldások egyik népszerű példája az Apache Airflow, amely lehetővé teszi az összetett számítási folyamatok szervezését és az ETL-csatornák kezelését. Az Airflow mellett más hasonló megoldásokat is aktívan használnak, mint például az Apache NiFi az adatok továbbítására és áramlására, valamint az n8n az üzleti folyamatok automatizálására.



FEJEZET 7.4.

AZ ETL ÉS MUNKAFOLYAMATOK ÖSSZEHANGOLÁSA: GYAKORLATI MEGOLDÁSOK

DAG és Apache Airflow: munkafolyamatok automatizálása és összehangolása

Az Apache Airflow egy ingyenes és nyílt forráskódú platform, amelyet a munkafolyamatok (ETL - conveyors) automatizálására, összehangolására és felügyeletére terveztek.

Nap mint nap nagy mennyiségű adattal kell dolgozni:

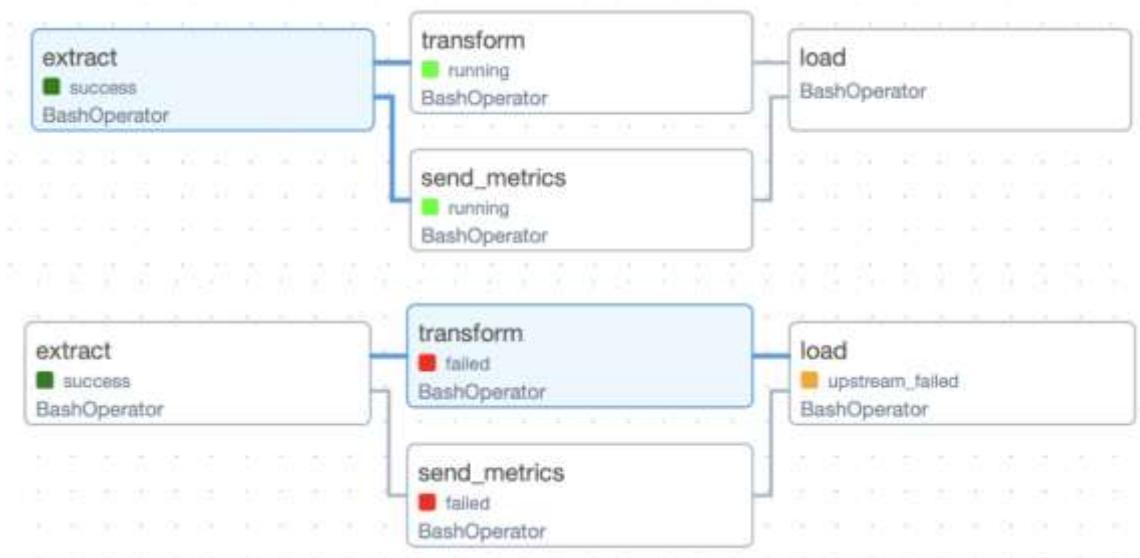
- Fájlok letöltése különböző forrásokból - kivonatolása (például beszállítótól vagy ügyfelektől).
- Az adatok átalakítása a kívánt formátumba - A átalakítása (strukturálás, tisztítás és érvényesítés).
- Az eredmények elküldése ellenőrzésre és jelentések készítése - betöltése (feltöltés a szükséges rendszerekbe, dokumentumokba, adatbázisokba vagy műszerfalakba).

Az ilyen ETL-folyamatok kézi végrehajtása jelentős időt vesz igénybe, és az emberi hiba kockázatához vezet. Az adatforrás megváltozása vagy az egyik lépés meghibásodása késedelmet és hibás eredményeket okozhat.

Az olyan automatizálási eszközök, mint az Apache Airflow, lehetővé teszik egy megbízható ETL -szállító létrehozását, a hibák minimalizálását, a feldolgozási idő csökkentését és annak biztosítását, hogy az adatok minden szakaszban helyesek legyenek. Az Apache Airflow középpontjában a DAG (Directed Acyclic Graph) koncepciója áll - egy olyan irányított aciklikus gráf, amelyben minden feladat (operátor) más függőségekhez kapcsolódik, és szigorúan meghatározott sorrendben hajtják végre. A DAG kiküszöböli a ciklusokat, ami logikus és kiszámítható struktúrát biztosít a feladatok végrehajtásához.

Az Airflow gondoskodik az összehangolásról - a feladatok közötti függőségek kezeléséről, a végrehajtási ütemezések ellenőrzéséről, az állapot nyomon követéséről és a hibákra való automatikus reagálásról. Ez a megközelítés minimalizálja a kézi beavatkozást, és biztosítja a teljes folyamat megbízhatóságát.

A Task Orchestrator olyan eszköz vagy rendszer, amelyet a feladatok végrehajtásának kezelésére és ellenőrzésére terveztek komplex számítástechnikai és információs környezetekben. Megkönnyíti a feladatok végrehajtásának telepítését, automatizálását és irányítását a teljesítmény javítása és az erőforrások optimalizálása érdekében.



7.4-1. ábra Az Apache Airflow egy felhasznábarát felületet biztosít, ahol megjelenítheti a DAG -ETL -ETL -ETL-t, megtekintheti a végrehajtási naplókat, a feladatok indítási állapotát és még sok másit.

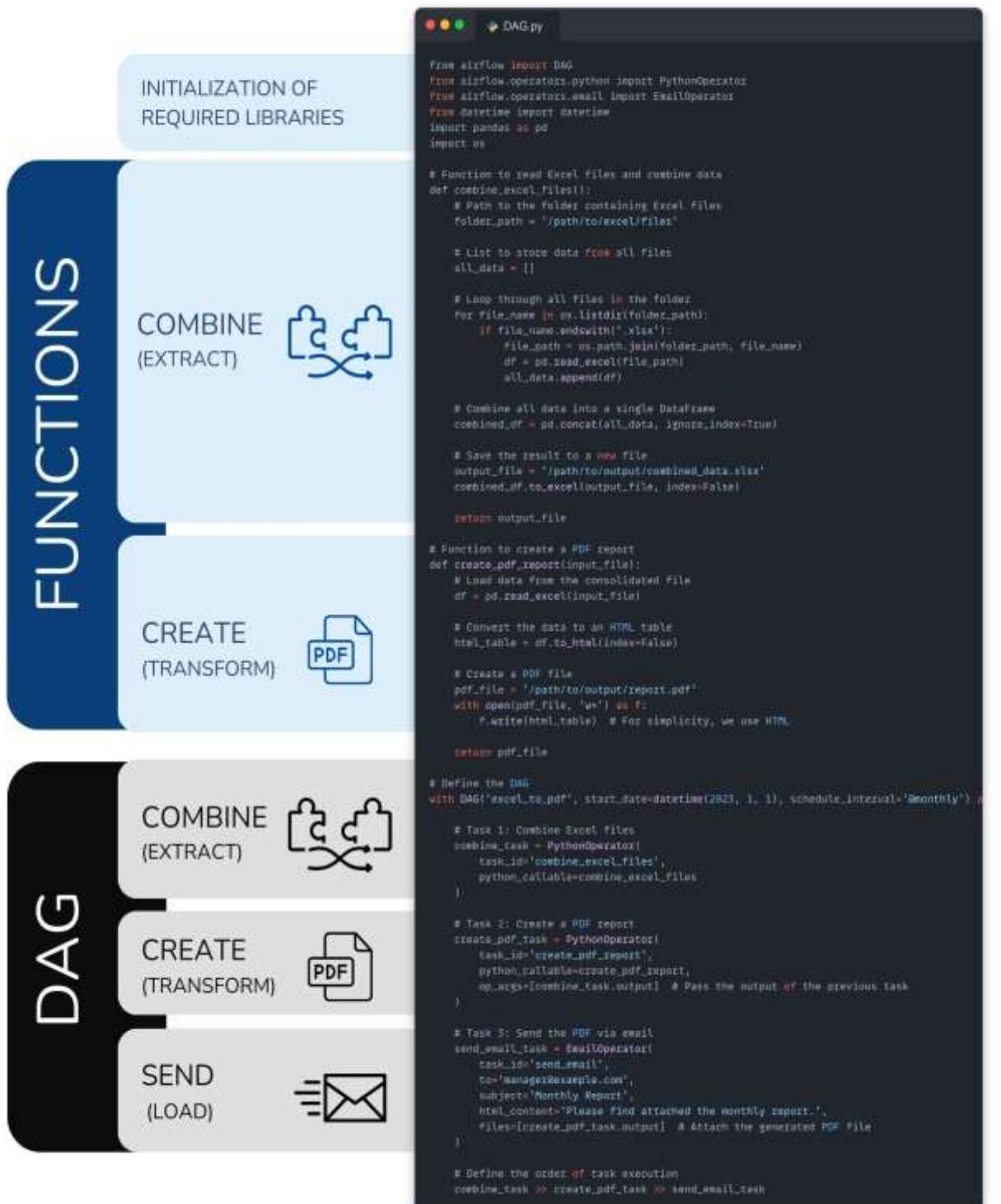
Az Airflowt széles körben használják az elosztott számítástechnika, az adatfeldolgozás, az ETL (Extract, Transform, Load) folyamatirányítás, a feladatütemezés és más adatforgatókönyvek összehangolására és automatizálására. Alapértelmezés szerint az Apache Airflow SQLite adatbázist használ.

Egy példa egy egyszerű DAG-ra, hasonlóan az ETL, a következő feladatokból áll: Extract, Transform és Load. A felhasználói felületen keresztül vezérelt gráfban (7.4-1. ábra) a feladatok (kód részletek) végrehajtási sorrendjét határozzuk meg: például először az extract, majd a transform (és a sending_metrics) végrehajtásra kerül, és a load feladat fejezi be a munkát. Ha minden feladat befejeződött, az adatbetöltési folyamat sikeresnek tekinthető.

Apache Airflow: gyakorlati alkalmazás az ETL automatizálására

Az Apache Airflow széles körben használatos az összetett adatfeldolgozási folyamatok szervezésére, lehetővé téve rugalmas ETL-szállítók építését. Az Apache Airflow webes felületen keresztül vagy programozottan, Python kódon keresztül futtatható (7.4-2. ábra). A webes felületen (7.4-3. ábra) a rendszerelők és a fejlesztők vizuálisan figyelemmel kísérhetik a DAG-okat, futtathatják a feladatokat és elemezhetik a végrehajtás eredményeit.

A DAG segítségével a feladatok egyértelmű sorrendjét állíthatja be, kezelheti a feladatok közötti függőségeket, és automatikusan reagálhat a forrásadatok változásaira. Nézzünk egy példát az Airflow használatára a jelentések feldolgozásának automatizálására (7.4-2. ábra).



Ábra 7.4-2 ETL -szállító koncepció az adatfeldolgozáshoz az Apache Airflow segítségével.

Ez a példa (7.4-2. ábra) a DAG-ot veszi figyelembe, amely az ETL -szállítóban kulcsfontosságú feladatokat

lát el:

Excel -fájlok olvasása (Extract):

- Az adott könyvtárban található összes fájl szekvenciális átnézése.
- Az adatok beolvasása minden egyes fájlból a pandas könyvtár segítségével.
- Az összes adat egyesítése egyetlen DataFrame-be.

PDF létrehozása -dokumentum (Transform):

- Az összevonot DataFrame átalakítása egy HTML -táblává.
- Mentse a táblázatot PDF formátumban (a demó verzióban - HTML-en keresztül).

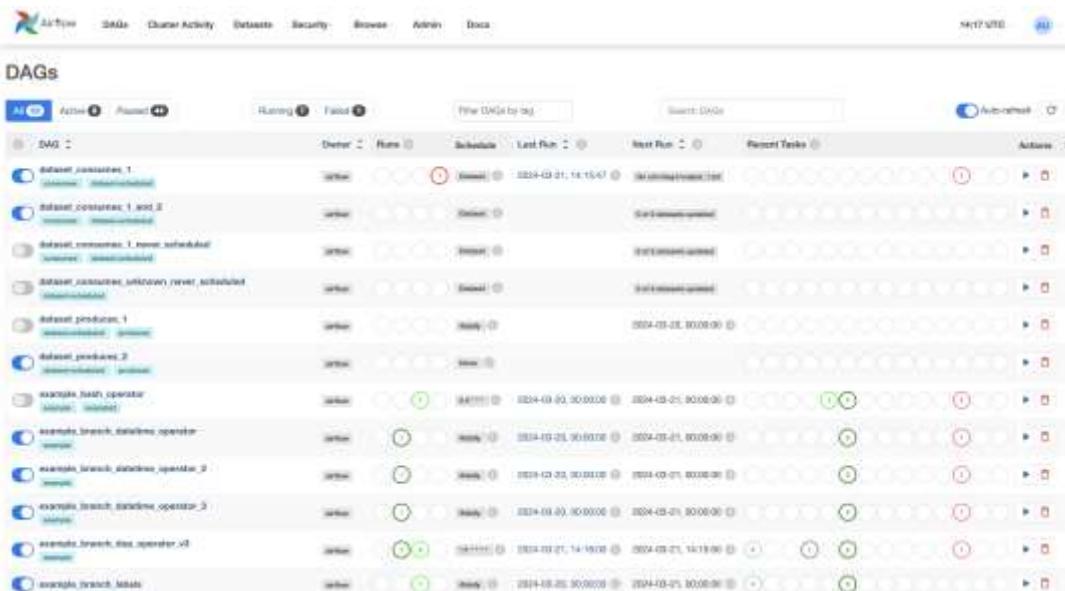
Jelentés küldése e-mailben (Load):

- EmailOperator alkalmazása a PDF -dokumentum e-mailben történő elküldéséhez.

DAG konfigurálása:

- A feladatok sorrendjének meghatározása: adatok kinyerése → jelentés készítése → küldése.
- Indítási ütemezés hozzárendelése (@havonta - minden hónap első napja).

Az automatizált ETL -példa (7.4-2. ábra) azt mutatja be, hogyan lehet adatokat gyűjteni az Excel -fájlokból, létrehozni egy PDF -dokumentumot, és elküldeni e-mailben. Ez csak egy az Airflow számos lehetséges felhasználási esete közül. Ez a példa bármilyen konkrét feladathoz igazítható az adatfeldolgozás egyszerűsítése és automatizálása érdekében.



Ábra 7.4-3 A környezetben található összes DAG áttekintése a legutóbbi futtatásokra vonatkozó információkkal.

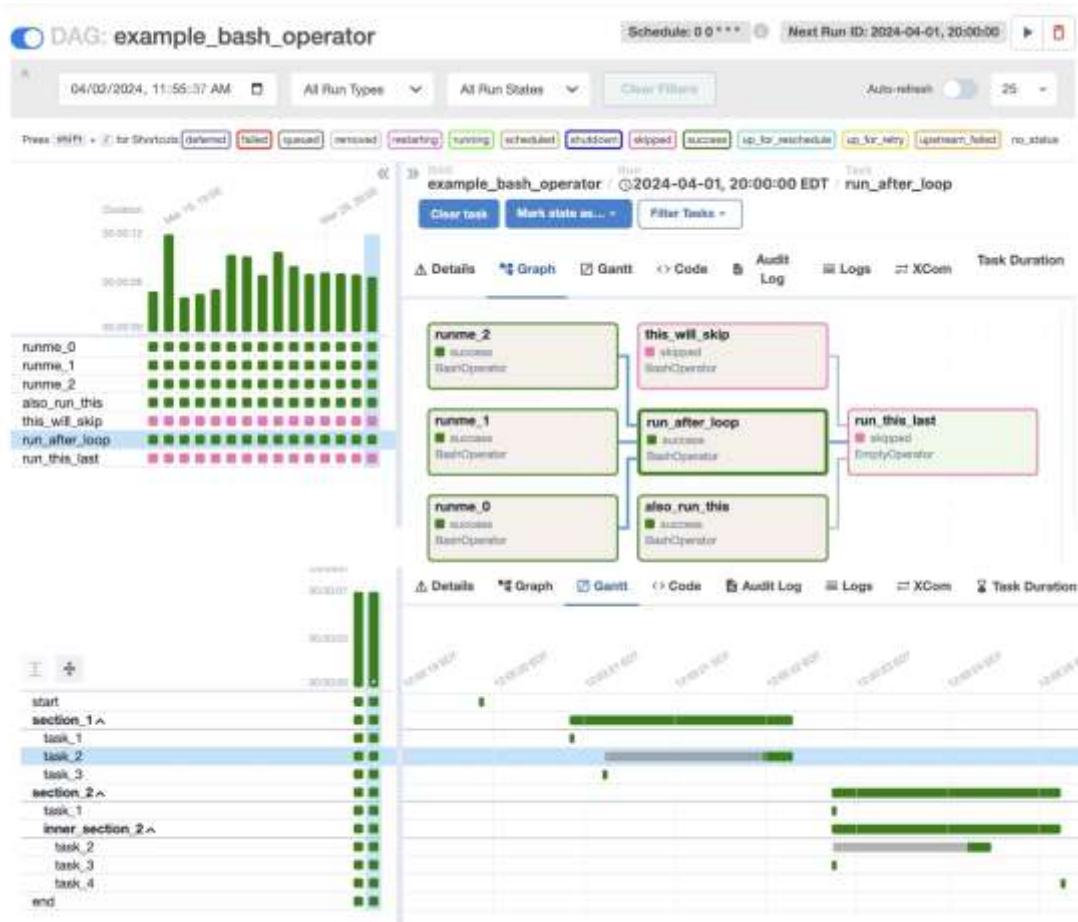
Az Apache Airflow webes felülete (7.4-3. ábra) átfogó vizuális környezetet biztosít az adatfolyamatok kezeléséhez. A DAG-okat interaktív gráfiként jeleníti meg, ahol a csomópontok a feladatokat, az élek pedig a köztük lévő függőségeket jelölik, így könnyen nyomon követhetők az összetett adatfolyamatok. A felület egy műszerfalat tartalmaz, amely a feladatok végrehajtásának állapotáról, a futtatási előzményekről, a részletes naplóról és a teljesítménymutatókról nyújt információt. A rendszergazdák egy intuitív felhasználói felületen keresztül manuálisan indíthatják el a feladatokat, indíthatják újra a sikertelen

műveleteket, függeszthetik fel a DAG-okat, és konfigurálhatják a környezeti változókat.

Az ilyen architektúra kiegészíthető adatérvényesítéssel, a végrehajtás állapotáról szóló értesítésekkel, valamint külső API-kkal vagy adatbázisokkal való integrációval. Az Airflow lehetővé teszi a DAG rugalmas adaptálását: új feladatok hozzáadása, sorrendjük megváltoztatása, láncok kombinálása - ami hatékony eszközzé teszi a komplex adatfeldolgozási folyamatok automatizálására. A DAG futtatásakor az Airflow webes felületén (7.4-3. ábra, 7.4-4. ábra) figyelemmel kísérheti a feladatok állapotát. A rendszer színes jelzést használ:

- Zöld - a feladat sikeresen végrehajtásra került.
- Sárga - a folyamat folyamatban van.
- Piros - hiba a feladat végrehajtása közben.

Hiba esetén (pl. hiányzó fájl vagy törött adatszerkezet) a rendszer automatikusan értesítést küld.



7.4-4. ábra Az Apache Airflow jelentősen leegyszerűsíti a problémadiagnosztikát, a folyamatoptimalizálást és a csoportos együttműködést az összetett adatfeldolgozási csővezetékeken.

Az Apache Airflow azért kényelmes, mert automatizálja a rutinfeladatokat, így nem kell azokat manuálisan elvégezni. A folyamatok végrehajtásának figyelemmel kísérésével és azonnali hibaüzenetekkel biztosítja a megbízhatóságot. A rendszer rugalmassága megkönnyíti új feladatok hozzáadását vagy a meglévők módosítását, így a munkafolyamatok a változó követelményekhez igazodnak.

Az Apache Airflow mellett léteznek hasonló eszközök a munkafolyamatok összehangolására. A nyílt forráskódú és ingyenes Prefect (7.3-5. ábra) például egyszerűbb szintaxist kínál, és jobban integrálódik a Pythonba, a Spotify által kifejlesztett Luigi hasonló funkciókat biztosít, és jól működik nagy adatokkal. Érdemes még megemlíteni a Kronost és a Dagstert, amelyek modern megközelítést kínálnak a Pipeline építéséhez, a modularitásra és a skálázhatóságra összpontosítva. A feladat-orchestrálási eszköz kiválasztása a projekt konkrét igényeitől függ, de mindegyik segít az összetett ETL adatfolyamatok automatizálásában.

Külön említést érdemel az Apache NiFi, egy nyílt forráskódú platform, amelyet adatfolyam- és útválasztási feladatokra terveztek. Az Airflow-val ellentétben, amely a kötegelt feldolgozásra és a függőségek kezelésére összpontosít, a NiFi a valós idejű, menet közbeni adatátalakításra és a rendszerek közötti rugalmas útválasztásra összpontosít.

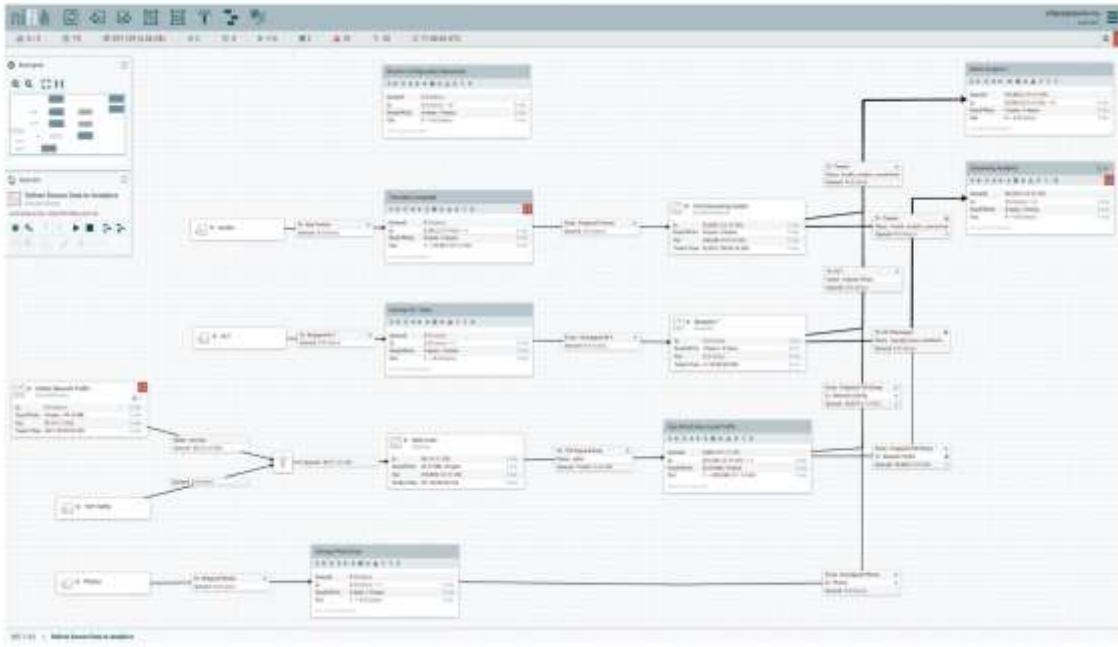
Apache NiFi útválasztáshoz és adatkonverzióhoz

Az Apache NiFi egy nagy teljesítményű nyílt forráskódú platform, amelyet a különböző rendszerek közötti adatáramlás automatizálására terveztek. Eredetileg 2006-ban az amerikai Nemzetbiztonsági Ügynökség (NSA) fejlesztette ki "Niagara Files" néven belső használatra. A projektet 2014-ben nyílt forráskódúvá tették és átadták az Apache Software Foundationnek, amely így a technológiaátadási kezdeményezések részévé vált [141].

Az Apache NiFi az adatok valós idejű gyűjtésére, feldolgozására és továbbítására szolgál. Az Airflow-val ellentétben, amely kötegelt feladatokkal dolgozik, és jól meghatározott ütemezést igényel, a NiFi folyamfeldolgozási üzemmódban működik, lehetővé téve az adatok folyamatos továbbítását a különböző szolgáltatások között.

Az Apache NiFi ideális az IoT eszközökkel, építési érzékelőkkel, felügyeleti rendszerekkel való integrációhoz, valamint például a CAD formátumok szerveren történő streaming validálásához, ahol azonnali reagálásra lehet szükség az adatok változásaira.

A beépített szűrési, transzformációs és útválasztási eszközökkel a NiFi lehetővé teszi az adatok szabványosítását (Transform), mielőtt azokat továbbítanák (Load) a tároló- vagy elemzőrendszerekbe. Az egyik fő előnye a beépített biztonsági támogatás és hozzáférés-szabályozás, ami megbízható megoldást jelent az érzékeny információk kezelésére.



7.4-5. ábra Az adatáramlás grafikus ábrázolása az Apache NiFi felületén.

Az Apache NiFi hatékonyan kezeli a valós idejű adatáramlási, szűrési és útválasztási feladatokat. Ideális olyan technikailag intenzív forgatókönyvekhez, ahol fontos a stabil információátvitel a rendszerek között és a nagy áteresztképesség.

Ha azonban a fő cél a különböző szolgáltatások integrálása, a rutinműveletek automatizálása és a munkafolyamatok gyors beállítása mély programozási ismeretek nélkül, akkor az alacsony belépési küszöbbel és maximális rugalmassággal rendelkező megoldásokra van szükség. Az egyik ilyen eszköz az n8n - egy Low-Code /No-Code osztályú platform, amely az üzleti automatizálásra és a folyamatok vizuális hangszerelésére összpontosít.

n8n Alacsony kódú, No-Code folyamat-orchestrálás

Az **n8n** egy nyílt forráskódú, alacsony kódú / No-Code platform automatizált munkafolyamatok létrehozására, amelyet az egyszerű használat, a rugalmasság és a külső szolgáltatások széles körével való gyors integrálhatóság jellemzi.

A No-Code egy olyan módszer, amellyel digitális termékeket lehet létrehozni kódírás nélkül. A folyamat minden eleme - a logikától a felületig - kizárolag vizuális eszközök segítségével valósul meg. A No-Code platformok a technikai háttérrel nem rendelkező felhasználóknak szólnak, és lehetővé teszik automatizációk, űrlapok, integrációk és webes alkalmazások gyors létrehozását. Példa: egy felhasználó programozási ismeretek nélkül, drag-and-drop felületen keresztül állíthatja be az automatikus értesítés-küldést vagy a Google Sheets integrációt.

A nyílt forráskódú és helyi telepítési képességekkel az n8n az automatizálási és ETL Pipelines létrehozásá

folyamatokban teljes ellenőrzést biztosít a vállalatok számára az adataik felett, miközben biztosítja a biztonságot és a felhőszolgáltatóktól való függetlenséget.

Ellentétben az Apache Airflow, amely a merev szervezésű számítási feladatokra irányul, és Python ismeretet igényel, az n8n vizuális szerkesztőt biztosít, amely lehetővé teszi a szkriptek írását anélkül, hogy programozási nyelvek ismeretére lenne szükség (7.4-6. ábra). Bár felülete lehetővé teszi az automatizált folyamatok létrehozását kódírás nélkül (No-Code), összetettebb forgatókönyvek esetén a felhasználók saját JavaScript és Python -funkciókkal bővíthetik a képességeket (Low-Code).

A Low-Code a szoftverfejlesztés olyan megközelítése, amelyben egy alkalmazás vagy folyamat alapvető logikája grafikus felület és vizuális elemek segítségével jön létre, és a programkódot csak a funkciók testreszabására vagy bővítésére használják. A Low-Code platformok lehetővé teszik a megoldások fejlesztésének jelentős felgyorsítását azáltal, hogy nem csak programozókat, hanem alapvető műszaki ismeretekkel rendelkező üzleti felhasználókat is bevonnak. Példa: egy felhasználó kész blokkokból építhet fel egy üzleti folyamatot, és szükség esetén hozzáadhatja saját szkriptjét JavaScript vagy Python nyelven.

Bár az n8n alacsony belépési küszöbbel rendelkező platformként van pozicionálva, alapvető programozási ismeretek, a webes technológiák megértése és az API. A rendszer rugalmassága lehetővé teszi, hogy a feladatok széles köréhez igazodjon - az automatizált adatfeldolgozástól az üzenetküldőkkel, IoT -eszközökkel és felhőszolgáltatásokkal való integrációjig.

Az n8n legfontosabb jellemzői és előnyei:

- **A nyílt forráskódú** és helyi telepítési lehetőségek teljes körű adatellenőrzést, biztonsági megfelelőséget és a felhőszolgáltatóktól való függetlenséget biztosítanak.
- **Integráció több mint 330 szolgáltatással**, beleértve a CRM, az ERP, az e-kereskedelem, a felhőplatformok, az üzenetküldők és az adatbázisok **integráját**.
- **A forgatókönyvek rugalmassága**: az egyszerű értesítésekkel az összetett láncokig, a -kérelmek API-feldolgozásával, a döntési logikával és az AI -szolgáltatások összekapcsolásával.
- **A JavaScript és Python támogatása**: a felhasználók szükség szerint beágyazhatnak egyéni kódot, így bővíve az automatizálási lehetőségeket.
- **Intuitív vizuális felület**: lehetővé teszi az összes folyamatlépés gyors konfigurálását és vizualizálását.

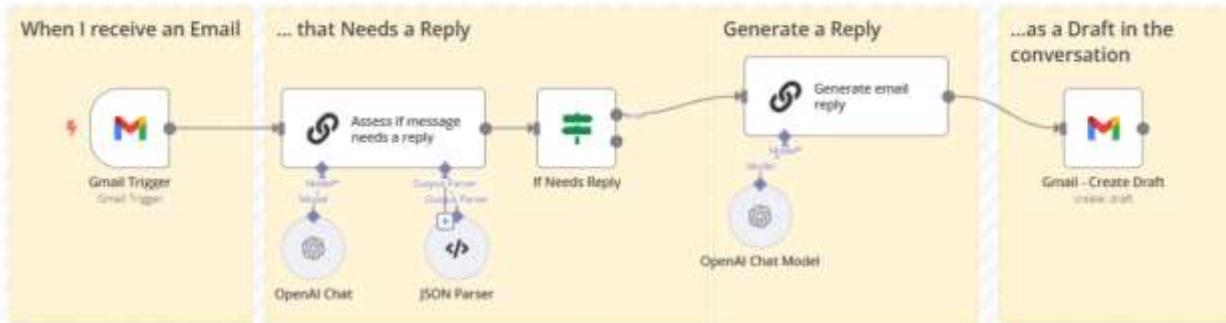
A Low-Code osztályú platformok olyan eszközöket biztosítanak, amelyekkel minimális kóddal hozhatók létre digitális megoldások, így ideálisak olyan csapatok számára, amelyek nem rendelkeznek mély műszaki szakértelemmel, de automatizálniuk kell a folyamatokat.

Az építőiparban az n8n számos folyamat automatizálására használható, például a projektmenedzsment rendszerekkel való integrációra, a folyamellenőrzésre, a jelentések és levelek készleten kívüli írására, az anyagkészlet-adatok automatikus frissítésére, a feladatok állapotáról szóló értesítések küldésére a csapatoknak és még sok másra. Az n8n-ben testre szabott Pipeline többszörösére csökkentheti a kézi műveleteket, csökkentheti a hibák valószínűségét és felgyorsíthatja a projektvégrehajtáshoz szükséges döntéshozatalt.

Közel kétezer kész, ingyenes és nyílt forráskódú n8n Pipeline, amely a következő címen érhető el: n8n.io/workflows, közül választhat az építési munkafolyamatok és a személyes feladatok automatizálására, a rutinműveletek csökkentésére.

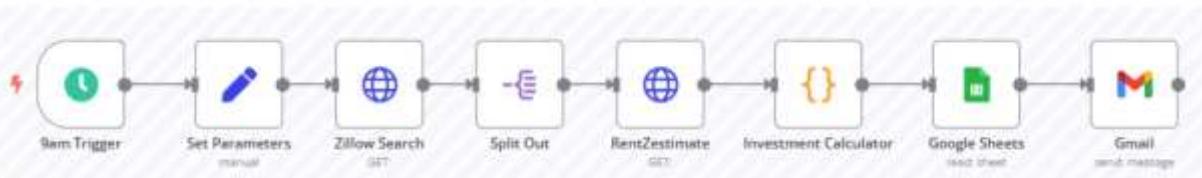
Vegyük például a, az n8n.io [142] oldalon ingyenesen elérhető kész Pipeline sablonok egyikét, amely automatikusan választervezeteket készít a Gmailben (7.4-6. ábra), segítve ezzel azokat a felhasználókat, akik nagy mennyiségű e-mailt kapnak, vagy nehezen tudnak válaszokat írni.

Ez az n8n "Gmail AI Auto-Responder: Választervezetek létrehozása bejövő e-mailekre" sablon (7.4-6. ábra) elemzi a bejövő e-maileket a ChatGPT LLM segítségével, megállapítja, hogy szükség van-e válaszra, létrehoz egy vázlatot a ChatGPT-ből, majd a szöveget HTML -re konvertálja, és hozzáadja a Gmail üzenetláncához. Ez nem küldi el automatikusan az e-mailt, lehetővé teszi a válasz kézi szerkesztését és jóváhagyását. A beállítás körülbelül 10 percet vesz igénybe, és magában foglalja a Gmail API és az OpenAI API integráció OAuth konfigurációját. Az eredmény egy kényelmes és ingyenes megoldás a rutinszerű e-mail kommunikáció automatizálására anélkül, hogy elveszítené az e-mailek tartalma feletti ellenőrzést.



Ábra 7.4-6 Automatizált e-mail válaszgenerálás folyamata az n8n használatával.

Egy másik példa az automatizálásra az n8n segítségével az ingatlanpiacra a nagyszerű ajánlatok megtalálása [143]. Az N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", a Zillow API segítségével gyűjti össze a megadott kritériumoknak megfelelő napi releváns ajánlatokat. Automatikusan kiszámítja a legfontosabb befektetési mérőszámokat (Cash on Cash ROI, havi cash flow, Down Payment), frissíti a Google Sheets és egy összefoglaló jelentést küld e-mailben (7.4-7. ábra), lehetővé téve a befektetők számára, hogy időt takarítsanak meg és gyorsan reagáljanak a legjobb ajánlatokra.



Ábra: 7.4-7 Az ingatlanok befektetési vonzerejének értékelésére szolgáló automatizált folyamat.

Rugalmassága és bővíthetősége az n8n értékes eszközévé teszi a vállalatok számára, amelyek digitálisan átalakulni és versenyképesebbé válni kívánnak a piacra viszonylag egyszerű és ingyenes nyílt forráskódú eszközökkel.

Az olyan eszközök, mint az Apache NiFi, az Airflow és az n8n az adatfeldolgozás három rétegeként képzelhetők el (7.4-8. ábra). A NiFi kezeli az adatáramlást, biztosítva annak átadását és átalakítását, az Airflow az adatok feldolgozási csővezetékekbe történő összevonásával hangszereli a feladatok végrehajtását, az n8n pedig automatizálja a külső szolgáltatásokkal való integrációt és kezeli az üzleti logikát.



	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Ábra 7.4-8 Az Apache Airflow, az Apache NiFi és az n8n a modern adatkezelési architektúra három egymást kiegészítő rétegének tekintethető.

Ezek az ingyenes és nyílt forráskódú eszközök együttesen az építőiparban az adat- és folyamatmenedzsment hatékony ökoszisztemájának példáját alkotják, lehetővé téve a vállalatok számára, hogy az információkat a döntéshozatal és a folyamatok automatizálása érdekében felhasználják.

Következő lépések: a manuális műveletekről az analitikán alapuló megoldásokra való áttérés

A mai építőipari vállalatok nagyfokú bizonytalansággal jellemezhető környezetben működnek: változó anyagárak, késedelmes szállítások, munkaerőhiány és szoros projektlezárások. Az analitikus műszerfalak, az ETL -szállítók és a BI-rendszerek használata segít a vállalatoknak a problémás területek gyors azonosításában, az erőforrások hatékonyságának értékelésében és a változások előrejelzésében mielőtt azok pénzügyi veszteségekhez vezetnének.

Összefoglalva ezt a részt, érdemes kiemelni a főbb gyakorlati lépéseket, amelyek segítségével a tárgyalt technológiákat alkalmazhatja a napi feladatai során:

- Adatvizualizációk és analitikus műszerfalak megvalósítása
 - A kulcsfontosságú teljesítménymutatók (KPI-k) nyomon követésére szolgáló műszerfalak létrehozásának elsajátítása () .
 - Használjon vizualizációs eszközöket az adataihoz (Power BI, Tableau, Matplotlib, Plotly).
- Az adatfeldolgozás automatizálása az ETL -folyamatok segítségével.
 - Automatikus adatgyűjtés beállítása különböző forrásokból (dokumentáció, táblázatok,

CAD) ETL-folyamatok segítségével.

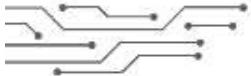
- Adattranszformáció megszervezése (pl. reguláris kifejezések ellenőrzése vagy számítás) Python szkriptek segítségével
- Próbálja ki az automatikus PDF (vagy DOC) jelentéskészítés beállítását az FPDF könyvtárral, az Excel fájlok ból származó adatok felhasználásával vagy más PDF dokumentumokból származó információk kinyerésével.

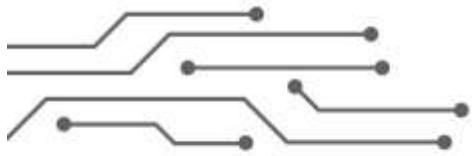
■ Nyelvi modellek használata (LLM) az automatizáláshoz

- Nagyméretű nyelvi modellek (LLM) használata kód generálásához, hogy segítsen adatokat kinyerni és elemezni strukturáltan dokumentumokból.
- Ismerkedjen meg az n8n automatizálási eszközével, és fedezze fel a kész sablonokat és esettanulmányokat a weboldalukon. Határozza meg, hogy az Ön munkájából mely folyamatok automatizálhatók teljes mértékben a No-Code/Low-Code megközelítéssel.

Az adatok és a folyamatok automatizálásának analitikus megközelítése nemcsak a rutinműveletekre fordított időt csökkenti, hanem javítja a döntéshozatal minőségét is. Azok a vállalatok, amelyek vizuális analitikai eszközöket alkalmaznak és ETL -szállító eszközöket, lehetőséget kapnak arra, hogy gyorsan reagáljanak a változásokra.

Az üzleti folyamatok automatizálása olyan eszközökkel, mint az n8n, az Airflow és a NiFi, csak az első lépés a digitális érettség felé. A következő lépés az automatizálás alapjául szolgáló adatok minőségi tárolása és kezelése. A 8. részben részletesen megvizsgáljuk, hogy az építőipari vállalatok hogyan építhetnek ki fenntartható adattárolási architektúrát, a dokumentumok és több formátumú fájlok káoszából a központosított tárolási és elemzési platformok felé haladva.





VIII RÉSZ

ADATTÁROLÁS ÉS -KEZELÉS AZ ÉPÍTŐIPARBAN

A 8. rész az építőiparban alkalmazott modern adattárolási és adatkezelési technológiákat vizsgálja. Elemzi a nagy mennyiségű információ kezelésére alkalmas hatékony formátumokat - az egyszerű CSV és XLSX formátumtól a nagyobb teljesítményű Apache Parquet és az ORC formátumig, részletes összehasonlítással ezek képességeiről és korlátairól. Az adattárházak (DWH), adattavak) és hibrid megoldásai (Data Lakehouse), valamint az adatkezelés) és az adatminimalizmus) elvei kerülnek tárgyalásra. Részletesen foglalkozunk az adatmocsár problémáival) és az információs rendszerekben kialakult káosz megelőzésére irányuló stratégiákkal. Bemutatásra kerülnek az adatokkal való munka új megközelítései, beleértve a vektoros adatbázisokat és azok alkalmazását az építésben a Bounding Box fogalmán keresztül. Ez a rész kitér a DataOps és a VectorOps módszertanokra is, mint az adatmunkafolyamatok szervezésének új szabványaira.

FEJEZET 8.1.

ADATINFRASTRUKTÚRA: A TÁROLÁSI FORMÁTUMOKTÓL A DIGITÁLIS ADATTÁRAKIG

Adatatomok: a hatékony információkezelés alapja

Az Univerzumban minden a legkisebb építőelemekből - atomokból és molekulákból - áll, és idővel minden élő és élettelen dolog elkerülhetetlenül visszatér ebbe a kezdeti állapotba. A természetben ez a folyamat elképesztő gyorsasággal megy végbe, amit megpróbálunk átvinni az ember által irányított folyamatokra.

Az erdőben minden élő szervezet végül tápláló anyaggá alakul át, amely új növények alapjául szolgál. Ezek a növények pedig táplálékká válnak az új előlények számára, amelyek ugyanazokból az atomokból állnak, amelyek évmilliókkal ezelőtt létrehozták az Univerzumot.

Az üzleti világban is fontos, hogy a komplex, többrétegű struktúrákat a legalapvetőbb, minimálisan feldolgozott egységekre bontsuk - hasonlóan a természetben található atomokhoz és molekulákhöz. Ez lehetővé teszi az adatatomok hatékony tárolását és kezelését, így azok gazdag, termékeny alappá válnak, amely az analitika és a döntések minőségének növekedése szempontjából kulcsfontosságú erőforrássá válik.



Ábra: 8.1-1 Az elemzés és a döntéshozatal az egyszer már feldolgozott és tárolt, újra felhasznált adatokon alapul.

A zenei kompozíciók hangjegyeiből állnak, amelyek összetett zeneműveket alkotnak, míg a szavak egy kezdetleges egységből, a betű-hangból jönnek létre. Legyen szó a természetről, a tudományról, a

gazdaságról, a művészetről vagy a technológiáról, a világ figyelemre méltó egységet és harmóniát mutat a pusztítás, a struktúra, a körforgás és a teremtés törekvéseben. Hasonlóképpen, a folyamatok a költségszámítási rendszerekben is apró, strukturált egységekre - erőforráselemekre - bomlanak le a költségszámítás és az ütemtervek szintjén. Ezeket az egységeket, akárcsak a jegyzeteket, azután összetettebb számítások és ütemtervek kialakításához használják fel. Ugyanezt az elvet alkalmazzák a számítógéppel segített tervezési rendszerek is, amelyekben az összetett építészeti és mérnöki projekteket alapelemekből - egyedi elemekből és könyvtári komponensekből - építik fel, amelyekből egy komplex épület vagy szerkezet projektjének teljes 3D -modellje jön létre.

A természetben és a tudományban rejlő ciklikusság és struktúra fogalma a modern adatvilágban is tükröződik. Ahogy a természetben minden élőlény visszatér az atomokhoz és molekulákhoz, úgy a modern adatfeldolgozó eszközök világában az információ is hajlamos visszatérni a legprimitívebb formájához.

A véges oszthatatlanságukkal rendelkező legkisebb elemek az üzleti folyamatok alapvető építőkövei. Fontos, hogy már a kezdetektől fogva gondosan átgondoljuk, hogyan gyűjtsük össze, strukturáljuk (atomokra bontjuk) és tároljuk ezeket a különböző forrásokból származó apró építőelemeket. Az adatok rendszerezése és tárolása nem csupán az adatok alkotóelemeire való bontás kérdése. Ugyanilyen fontos annak biztosítása is, hogy az adatokat strukturált módon integrálják és tárolják, hogy az adatok könnyen visszakereshetők, elemezhetők és felhasználhatók legyenek a döntéshozatalhoz, amikor csak szükség van rájuk.

Az információk hatékony feldolgozásához gondosan meg kell választani az adattárolás formátumát és módszereit - ahogyan a talajt is elő kell készíteni a fák növekedéséhez. Az adattárolókat úgy kell megszervezni, hogy biztosítsák az információk magas minőségét és relevanciáját, kiküszöböльve a felesleges vagy irreleváns adatokat. Minél jobban strukturált ez az "információs talaj", annál gyorsabban és pontosabban találják meg a felhasználók a megfelelő adatokat és annál pontosabban oldják meg az elemzési problémákat.

Információ tárolása: fájlok vagy adatok

Az adattárházak lehetővé teszik a vállalatok számára, hogy összegyűjtsék és egyesítsék a különböző rendszerekből származó információkat, és így egyetlen központot hozzanak létre a további elemzésekhez. Az összegyűjtött múltbeli adatok nemcsak a folyamatok mélyebb elemzését teszik lehetővé, hanem az üzleti teljesítményt befolyásoló minták azonosítását is.

Tegyük fel, hogy egy vállalat egyszerre több projekten dolgozik. Egy mérnök szeretné megérteni, hogy mennyi betont öntöttek már ki, és mennyit kell még beszerezni. Hagyományos megközelítésben manuálisan kellene keresnie a szerveren, és több becslési táblázatot megnyitnia, összehasonlítnia az elvégzett munkák igazolásával, és ellenőriznie az aktuális készletegyenlegeket. Ez óráig vagy akár napokig is eltarthat. Még az ETL-folyamatok és az automatikus szkriptek mellett is félig manuális marad a feladat: a mérnöknek még mindig kézzel kell megadnia a szerveren lévő mappák vagy konkrét fájlok elérési útvonalát. Ez csökkenti az automatizálás általános hatását, mivel továbbra is értékes munkaidőt vesz el.

Az adatkezelésre való áttérés során a szerver fájlrendszerével való munka helyett a mérnök egy egységes tárolási struktúrához kap hozzáférést, ahol az információk valós időben frissülnek. Egyetlen lekérdezés -

kód, SQL -lekérdezés vagy akár egy LLM -ügynök hívása formájában - azonnal pontos adatokat szolgáltathat az aktuális egyenlegekről, az elvégzett munkamennyiségekről és a közelgő szállításokról, ha az adatokat előre előkészítették és egy olyan adattárházban egyesítették, ahol nem kell mappákban bolyongani, fájlok tucatjait megnyitni és manuálisan összehasonlítni az értékeket.

Az építőipari vállalatok sokáig PDF -dokumentumokat, DWG -rajzokat, RVT -modelleket és több száz és ezer Excel -táblázatot és más különböző formátumokat használtak, amelyeket a vállalat szerverein meghatározott mappákban tároltak, ami megnehezítette az információk keresését, ellenőrzését és elemzését. Ennek eredményeképpen a projektek befejezése után megmaradt fájlokat leggyakrabban visszahelyezik a szerverre, archív tároló mappákba, amelyeket a jövőben gyakorlatilag nem használnak. Az ilyen hagyományos fájlalapú adattárolás az adatáramlás növekedésével veszít jelentőségből, mivel az emberi hibáknak való kiszolgáltatottsága miatt.

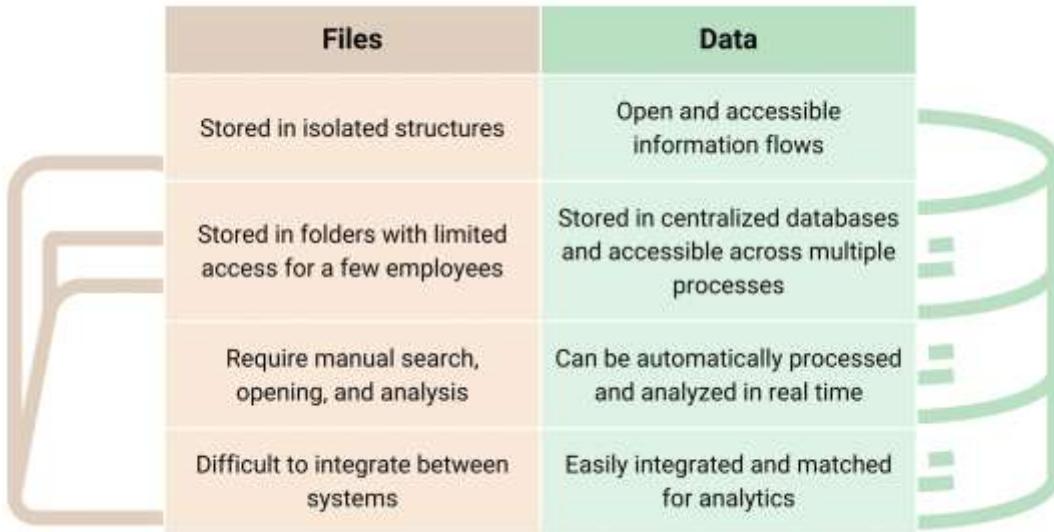
A fájl csak egy elszigetelt tároló, amelyben az adatokat tárolják. A fájlokat emberek, nem pedig rendszerek számára hozzák létre, ezért manuális megnyitást, olvasást és értelmezést igényelnek. Ilyen például egy Excel -táblázat, egy PDF -dokumentum vagy egy CAD -rajz, amelyet kifejezetten egy adott eszközzel kell megnyitni a kívánt információ eléréséhez. Strukturált visszakeresés és feldolgozás nélkül a benne lévő információ kihasználatlanul marad.

Az adatok viszont gépileg olvasható információk, amelyeket automatikusan összekapcsolnak, frissítenek és elemeznek. Egyetlen adattárházban (pl. adatbázis, DWH vagy Data Lake) az információ táblák, rekordok és kapcsolatok formájában jelenik meg. Ez lehetővé teszi az egységes tárolást, az automatizált lekérdezéseket, az értékelemzéseket és a valós idejű jelentéstételt.

A fájlok helyett adatok használata (8.1-1. ábra) lehetővé teszi a kézi keresési folyamat kiküszöbölését és a feldolgozás egységesítését. Az ilyen megközelítést már megvalósító vállalatok versenyelőnyre tesznek szert az információkhöz való gyors hozzáférés és az információk üzleti folyamatokba való gyors integrálhatósága miatt.

A fájlok használatáról az adatokra való áttérés elkerülhetetlen változás, amely az építőipar jövőjét fogja alakítani.

Az építőiparban minden vállalat kulcsfontosságú választás előtt áll: továbbra is széttagolt fájlokkban és silókban tárolja az információkat, amelyeket embereknek kell speciális programokkal olvasniuk, vagy a feldolgozás első szakaszában strukturált adatokká alakítja át azokat, és így egyetlen integrált digitális alapot hoz létre az automatizált projektmenedzsmenthez.



8.1-1. ábra Az információáramlás fejlődése: az elszigetelt fájloktól az integrált adatokig.

Az információrobbanás miatt a fájlok tárolásának és feldolgozásának hagyományos módszerei egyre kevésbé hatékonyak. Az építőiparban, akárcsak más ágazatokban, már nem elegendő a különböző fájIFORMÁTUMÚ, széttagolt fájlmappákra vagy a nem összekapcsolt adatbázisokra támaszkodni.

A digitális korban versenyképesnek maradni kívánó vállalatok elkerülhetetlenül integrált digitális platformok felé mozdulnak el, nagy adatszolgáltatási technológiákat és automatizált elemzőrendszereket használnak.

A fájlalapú tárolásról az adatvezérelt munkafolyamatokra való áttérés az információkezelési megközelítések újragondolását és a központosított adattárakba való további integrálásra alkalmas formátumok tudatos kiválasztását teszi szükségessé. Ez a választás határozza meg, hogy az adatok minden hatékonyan feldolgozhatók, minden gyorsan hozzáférhetők és minden könnyen integrálhatók a vállalat digitális folyamataiba.

Nagy adattárolás: a népszerű formátumok és hatékonyságuk elemzése

A tárolási formátumok kulcsszerepet játszanak az analitikai infrastruktúra skálázhatóságában, megbízhatóságában és teljesítményében. Az adatok elemzéséhez és feldolgozásához - például szűréshez, csoportosításhoz és aggregáláshoz - példáinkban a Pandas DataFrame - egy népszerű struktúrát használtunk a RAM-ban lévő adatokkal való munkához.

A Pandas DataFrame azonban nem rendelkezik saját tárolási formátummal, így a feldolgozás befejezése után az adatokat valamelyik külső formátumba - leggyakrabban CSV vagy XLSX - exportáljuk. Ezek a táblázatos formátumok könnyen cserélhetők és kompatibilisek a legtöbb külső rendszerrel, de számos korlátozással rendelkeznek: alacsony tárolási hatékonyság, tömörítés hiánya és gyenge verziókezelési támogatás:

- **CSV** (Comma-Separated Values - vesszővel elválasztott értékek): egyszerű szöveges formátum, amelyet széles körben támogatnak a különböző platformok és eszközök. Könnyen használható, de nem támogatja az összetett adattípusokat és a tömörítést.

- **XLSX** (Excel Open XML Spreadsheet): A Microsoft Excel fájiformátum, amely támogatja az olyan kifinomult funkciókat, mint a képletek, diagramok és a formázás. Bár hasznos a kézi adatelemzéshez és a vizualizációhoz, nem optimalizált nagyméretű adatfeldolgozásra.

A népszerű táblázatos XLSX és CSV mellett számos népszerű formátum létezik a strukturált adatok hatékony tárolására (8.1-2. ábra), amelyek mindegyike egyedi előnyökkel rendelkezik az egyedi adattárolási és elemzési követelményektől függően:

- **Apache Parquet**: oszlopos adattárolási fájiformátum, amelyet adatelemző rendszerekben való használatra optimalizáltak. Hatékony adattömörítési és kódolási sémákat kínál, így ideális összetett adatstruktúrákhöz és nagy adatfeldolgozáshoz.
- **Apache ORC** (Optimised Row Columnar): A Parquet -hez hasonlóan az ORC nagyfokú tömörítést és hatékony adattárolást biztosít. Optimalizálva van a nehéz olvasási műveletekre, és jól alkalmazható adattavak tárolására.
- **JSON** (JavaScript Object Notation): bár a JSON nem olyan hatékony az adattárolás szempontjából, mint az olyan bináris formátumok, mint a Parquet vagy az ORC, nagyon könnyen hozzáférhető és könnyen kezelhető, így ideális olyan szkriptek számára, ahol fontos az olvashatóság és a webes kompatibilitás.
- **Feather**: egy gyors, könnyű és könnyen használható, analízisorientált bináris oszlopos adattárolási formátum. Úgy terveztek, hogy hatékonyan továbbítsa az adatokat a Python (Pandas) és az R között, így kiváló választás az ezeket a programozási környezeteket érintő projekteket.
- **HDF5** (Hierarchical Data Format version 5): nagy mennyiségű adat tárolására és rendszerezésére terveztek. Az adattípusok széles skáláját támogatja, és jól alkalmazható összetett adatgyűjteményekkel való munkához. A HDF5 különösen népszerű a tudományos számítástechnikában, mivel képes a nagy adathalmazok hatékony tárolására és elérésére.

		XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame
	Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
	Compression	Built-in	None	High	Built-in	None (in-memory)
	Performance	Low	Medium	High	High	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Medium	Medium	Low
	Data Type Support	Limited	Very limited	Extended	Extended	Extended
	Scalability	Low	Low	High	High	Medium (memory limited)

8.1-2. ábra Az adatformátumok összehasonlítása, amely a tárolási és feldolgozási szempontok fő különbségeit mutatja.

Az ETL-folyamat betöltési szakaszában használt formátumok összehasonlító elemzésének elvégzése érdekében elkészült egy táblázat, amely a fájlméreteket és az olvasási időket mutatja (8.1-3. ábra). A vizsgálatban azonos adatokat tartalmazó fájlokat használtunk: a táblázat 10 000 sort és 10 véletlen értékekkel kitöltött oszlopot tartalmazott.

A tanulmány a következő tárolási formátumokat tartalmazza: CSV, Parquet, XLSX és HDF5, valamint ezek tömörített változatait ZIP-archívumokban. A nyers adatokat a NumPy könyvtár segítségével generáltuk, és Pandas DataFrame struktúraként ábrázoltuk. A vizsgálati folyamat a következő lépésekkel állt:

- Fájlménytés: az adatkeret négy különböző formátumban menthető: CSV, Parquet, XLSX és HDF5. Mindegyik formátum egyedi jellemzőkkel rendelkezik az adatok tárolásának módját illetően, ami hatással van a fájl méretére és az olvasási sebességre.
- ZIP fájltömörítés: a szabványos tömörítés hatékonyságának elemzése érdekében minden fájt tovább tömörítettünk egy ZIP-archívumba.
- Fájlolvasás (ETL - Load): az olvasási időt minden egyes fájl esetében a ZIP-ből való kicsomagolás után mértük. Ez lehetővé teszi az adathozzáférés sebességének becslését az archívumból való kicsomagolás után.

Fontos megjegyezni, hogy a Pandas DataFrame-et nem használtuk közvetlenül a méret vagy az olvasási idő elemzéséhez, mivel az nem egy önálló tárolási formátum. Csak a köztes struktúrájaként szolgált az adatok generálásához és későbbi, különböző formátumokban történő mentéséhez.



8.1-3. ábra A tárolóformátumok összehasonlítása méret és olvasási sebesség szerint.

A CSV és a HDF5 fájlok (8.1-3. ábra) nagy tömörítési hatékonyságot mutatnak, jelentősen csökkentve méretüket, amikor ZIP-be csomagolják őket, ami különösen hasznos lehet a tárolás optimalizálását igénylő forgatókönyvekben. Az XLSX fájlok ezzel szemben gyakorlatilag tömöríthetetlenek, és méretük ZIP-ben az eredetihez hasonló marad, ami kevésbé kedvezővé teszi őket nagy adatmennyiségek vagy olyan környezetek esetében, ahol az adathozzáférés sebessége fontos. Ezenkívül az XLSX olvasási ideje jelentősen magasabb a többi formátumhoz képest, így kevésbé előnyös a gyors adatolvasási műveletekhez. Az Apache Parquet oszlopos szerkezetének köszönhetően nagy teljesítményt mutat az analitikai feladatok és a nagy adatmennyiségek esetében.

Tárolás optimalizálása az Apache Parquet segítségével

A nagyméretű adatok tárolására és feldolgozására szolgáló egyik népszerű formátum az Apache Parquet. Ezt a formátumot kifejezetten oszlopos tárolásra terveztek (hasonlóan a Pandashoz), ami lehetővé teszi a memóriaigény jelentős csökkentését és az analitikus lekérdezések sebességének növelését. A hagyományos formátumokkal, például a CSV és az XLSX formátummal ellentétben a Parquet támogatja a natív tömörítést, és a nagy adatrendszerkhez, köztük a Sparkhoz, a Hadoophoz és a felhőalapú tároláshoz optimalizált.

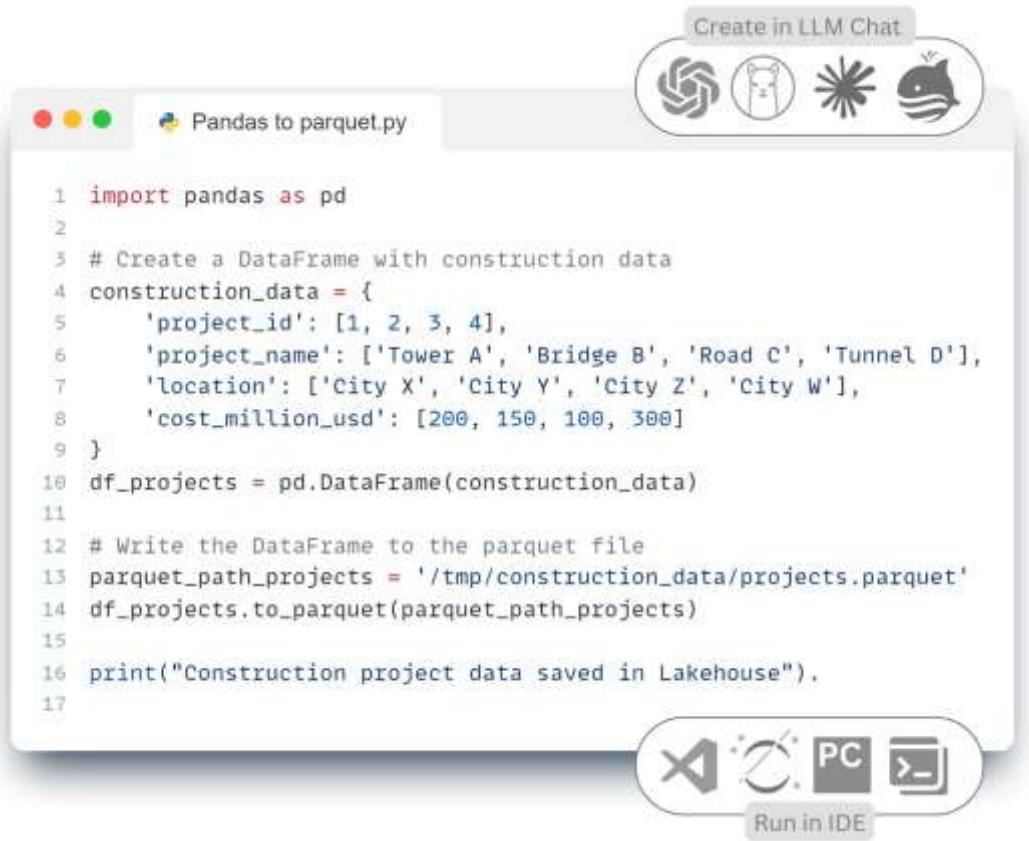
A Parquet fő jellemzői közé tartozik az adattömörítés és kódolás támogatása, amely jelentősen csökkenti a tárolási méretet és felgyorsítja az adatolvasási műveleteket, mivel az adatok összes sora helyett közvetlenül a kívánt oszlopokon dolgozik.

Egy szemléltető példát mutatunk arra, hogy milyen egyszerű az adatok Apache Parquet -ra való konvertálásához szükséges kódot megszerezni, használjuk az LLM -et.

- ❷ Küldjön szöveges kérést az LLM chat (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Írj kódot az adatok mentésére Pandas DataFrame Apache Parquet. ↵

- ❸ LLM válasza :



```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

The screenshot shows a code editor window titled "Pandas to parquet.py". The code itself is a Python script that imports pandas, creates a dictionary "construction_data" with four projects, and then converts it into a DataFrame named "df_projects". This DataFrame is then written to a parquet file at the path "/tmp/construction_data/projects.parquet". Finally, a print statement is included to confirm the save. At the bottom of the code editor, there are icons for running the code in an IDE, a terminal, or a browser, along with a "Run in IDE" button.

A 8.1-4 ábra Dataframe adatok átvitele a RAM-ból egy tárolási szempontból hatékony Apache Parquet formátumba néhány sor Python segítségével.

Következő példa: szimuláljuk az ETL-folyamatot a Parquet formátumban tárolt adatokkal, hogy a projekteket a "cost_million_usd" attribútumok egyikének bizonyos értéke alapján szűrjük (8.1-4. ábra).

- ❸ Miközben a csevegés folytatódik, küldjön egy szöveges kérést az LLM címre:

Írunk kódot, amelyben egy táblázat adatait szeretnénk szűrni, és csak azokat a projekteket (táblázat sorait) menteni az Apache Parquet adataiból, amelyek költsége (cost_million_usd paraméter) meghaladja a 150 millió dollárt. ↴

❷ LLM válasza :



The screenshot shows a user interface for generating code. At the top right is a button labeled "Create in LLM Chat". Below it are four circular icons representing different AI models or functions. The main area is a code editor window titled "ETL parquet.py" containing the following Python code:

```

1 # Loading construction project data
2 df_loaded_projects = pd.read_parquet(parquet_path_projects)
3
4 # Filtering out projects with a value greater than $150 million
5 df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]
6
7 # Saving the filtered data to a new Parquet file
8 filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
9 df_filtered_projects.to_parquet(filtered_parquet_path_projects)
10
11 print("Filtered construction projects data saved to Lakehouse.")

```

At the bottom right of the code editor is a button labeled "Run in IDE".

8.1-5. ábra Az ETL-folyamat az Apache Parquet formátumú adatokkal való munka során ugyanúgy néz ki, mint más strukturált formátumok esetében.

A Parquet formátum használata (az XLSX, CSV stb. formátummal szemben) a jelentősen csökkenti a tárolt információk mennyiségett és felgyorsítja a keresési műveleteket. Ezáltal kiválóan alkalmas az adatok tárolására és elemzésére egyaránt. A Parquet integrálható különböző feldolgozórendszerekkel, hatékony hozzáférést biztosítva hibrid architektúrákban.

A hatékony tárolási formátum azonban csak egy eleme a teljes adatélménynek. A fenntartható és skálázható környezet létrehozásához jól megtervezett adatkezelési architektúrára van szükség. Ezt a funkciót a DWH (Data Warehouse) osztályú rendszerek látják el. Ezek biztosítják a heterogén forrásokból származó adatok aggregálását, az üzleti folyamatok átláthatóságát, valamint a BI-eszközök és gépi tanulási algoritmusok segítségével történő komplex elemzés lehetőségét.

DWH: Adattárház adattárházak

Ahogy a Parquet formátumot nagy mennyiségű információ hatékony tárolására optimalizálták, úgy az Adattárház az adatok integrálására és strukturálására optimalizált az analitika, az előrejelzés és a vezetői döntéshozatal támogatása érdekében.

A mai vállalatoknál az adatok számos különböző forrásból származnak: ERP, CAFM, CPM, CRM rendszerek, könyvelés és raktárkezelés, épületek digitális CAD modelljei, IoT érzékelők és egyéb megoldások. Ahhoz, hogy holisztyikus képet kapunk, nem elég egyszerűen összegyűjteni az adatokat - azokat rendszerezni, szabványosítani és egyetlen tárolóhelyen központosítani kell. A DWH pontosan ezt teszi - egy

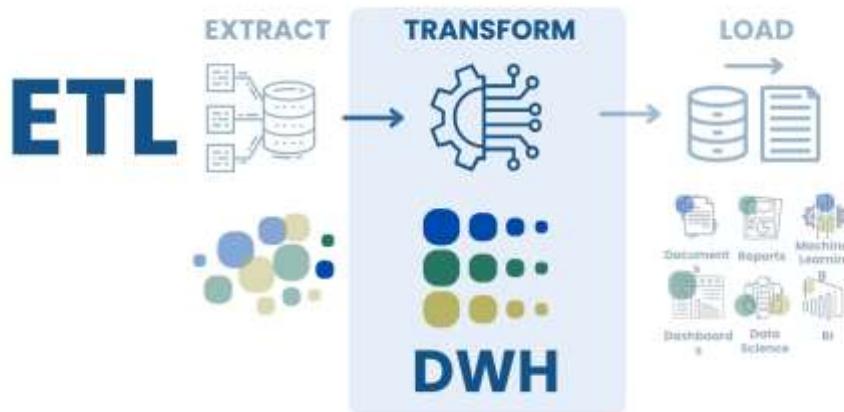
központosított tárolórendszer, amely lehetővé teszi a különböző forrásokból származó információk összegyűjtését, strukturálását és rendelkezésre bocsátását az elemzések és a stratégiai irányítás számára.

A DWH (Data Warehouse) egy olyan központosított adattárházrendszer, amely több forrásból származó információkat aggregál, strukturál, és elérhetővé teszi azokat az elemzésekhez és a jelentéstételhez.

Sok vállalatnál az adatok különböző rendszerekben vannak szétszórva, amelyeket a könyv első részeiben tárgyalunk (1.2-4. ábra). A DWH integrálja ezeket a forrásokat, biztosítva az információk teljes átláthatóságát és megbízhatóságát. A DWH adattárház egy olyan speciális adatbázis (egy nagy adatbázis), amely több forrásból származó adatokat gyűjt, dolgoz fel és tárol. A DWH fő jellemzői a következők:

- **Az ETL -folyamatok** (Extract, Transform, Load) **használata** - az adatok kinyerése a forrásokból, tisztítása, átalakítása, betöltése az adattárba, és ezen folyamatok automatizálása, amelyeket a könyv hetedik részében tárgyalunk.
- **Az adatok szemcsézettsége** - a DWH oldalon az adatok minden összesített formában (összefoglaló jelentések), minden szemcsézettségben (nyers adatok) tárolhatók. 2024-től a CAD- gyártók azok, akik granuláris adatokról beszélnek [125], ami talán azt jelzi, hogy az iparág felkészül a digitális épületmodell-adatok kezelésére specializált felhőalapú tárolásra való áttérésre.
- **Az analitikát támogató és prediktív** - adattárházak biztosítják a BI eszközök, a Big Data -elemzés és a gépi tanulás alapját.

A DWH az üzleti intelligencia alapjául szolgál, lehetővé téve a kulcsfontosságú teljesítménymutatók elemzését, az értékesítés, a beszerzések és a költségek előrejelzését, valamint az automatizált jelentéstételt és az adatok vizualizálását (8.1-6. ábra).



8.1-6 ábra Egy ETL -folyamatban a DWH központi tárolóként működhet, ahol a különböző rendszerekből kinyert adatok átalakítási és kirakodási szakaszokon mennek keresztül.

A DWH kulcsszerepet játszik az információk integrálásában, tisztításában és strukturálásában, szilárd alapot biztosítva az üzleti intelligencia és a döntéshozatali folyamatok számára. A mai környezetben azonban, ahol az adatmennyiség gyorsan növekszik és az adatforrások egyre sokszínűbbé válnak, az információtárolás hagyományos DWH megközelítése gyakran igényel bővítést ELT és Data Lake

formájában

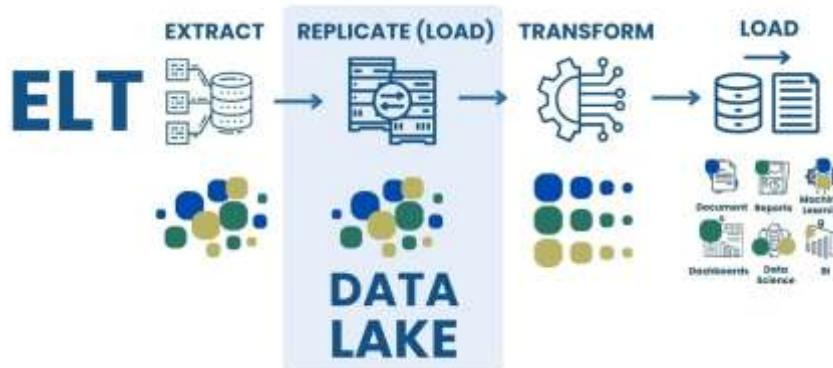
Data Lake - az ETL evolúciója az ELT: a hagyományos tisztítástól a rugalmas feldolgozásig.

A klasszikus DWH - adattárházak, amelyeket úgy terveztek, hogy strukturált adatokat tároljanak elemző lekérdezésekre optimalizált formátumban, korlátokba ütköztek a strukturálatlan adatok kezelése és a skálázhatóság terén. E kihívásokra válaszul jelentek meg az adattavak), amelyek nagy mennyiségi heterogén adat rugalmas tárolását kínálják.

A Data Lake egy alternatív DWH -megközelítést kínál, amely lehetővé teszi a strukturálatlan, félig strukturált és nyers adatokkal való munkát előzetes merev séma nélkül. Ez a tárolási módszer gyakran releváns a valós idejű adatfeldolgozás, a gépi tanulás és a fejlett analitika szempontjából. A DWH-val ellentétben, amely az adatokat betöltés előtt strukturálja és aggregálja, a Data Lake lehetővé teszi az információk nyers formában történő tárolását, így rugalmasságot és skálázhatóságot biztosít

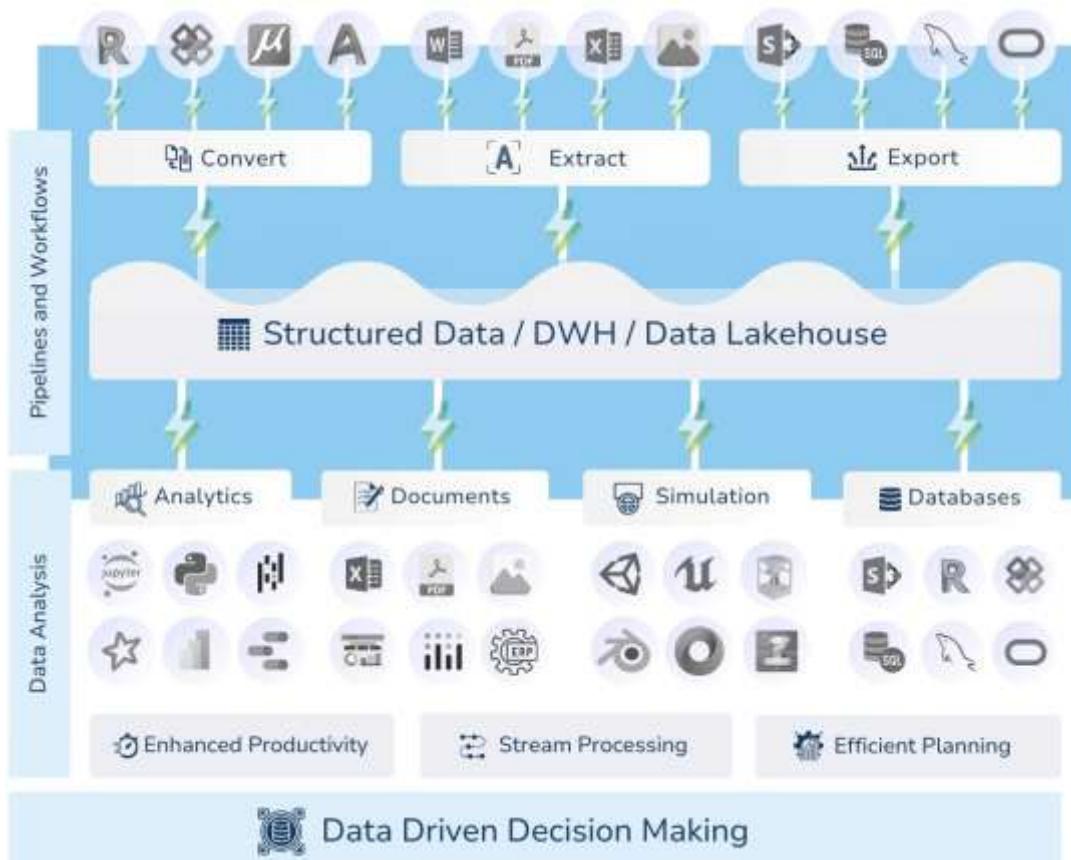
A hagyományos adattárházakkal (RDBMS, DWH) szembeni csalódottság és a "nagy adatok" iránti érdeklődés vezetett az adattavak megjelenéséhez, ahol a bonyolult ETL helyett az adatokat egyszerűen feltöltik egy lazán strukturált tárolóba, és a feldolgozás az elemzési szakaszban történik:

- A hagyományos adattárházakban az adatokat jellemzően előfeldolgozzák, átalakítják és megtisztítják (ETL - Extract, Transform, Load), mielőtt betöltenék a raktárba (8.1-6. ábra). Ez azt jelenti, hogy az adatokat strukturálják és optimalizálják a konkrét jövőbeli elemzési és jelentési feladatokhoz. A hangsúly a nagy lekérdezési teljesítmény és az adatintegritás fenntartásán van. Ez a megközelítés azonban költséges és kevésbé rugalmas lehet az új adattípusok és a gyorsan változó adatsémák integrálása szempontjából.
- Az adattavakat viszont úgy terveztek, hogy nagy mennyiségi nyers adatot tároljanak eredeti formátumban (8.1-7. ábra). Az ETL (Extract, Transform, Load), folyamatot felváltja az ELT (Extract, Load, Transform), ahol az adatokat először "úgy, ahogy vannak" betöltenek a raktárba, és csak ezután lehet azokat szükség szerint átalakítani és elemezni. Ez nagyobb rugalmasságot biztosít, és lehetővé teszi a heterogén adatok tárolását, beleértve a strukturálatlan adatokat, például a szöveget, a képeket és a naplókat.



Ábra 8.1-7 Az ETL -től eltérően a Data Lake az ELT -t használja, amelyben az információ először "nyers" formában kerül feltöltésre, és az átalakítás a feltöltési szakaszban történik.

A hagyományos adattárházak a nagy lekérdezési teljesítmény biztosítása érdekében az adatok előfeldolgozására összpontosítanak, míg az adattavak a rugalmasságot helyezik előtérbe: nyers adatokat tárolnak és szükség szerint átalakítják azokat (8.1-8. ábra).



8.1-8 A modern tárolási koncepciók célja, hogy minden típusú adatot tároljanak és feldolgozzanak döntéshozatali célokra.

Azonban minden előnye ellenére az adattavak és az adattavak nem hátrányok nélküliek. A szigorú struktúra hiánya és az információkezelés összetettsége káoszhöz vezethet, amelyben az adatok duplikálódnak,

ellentmondanak egymásnak vagy irrelevánssá válnak. Ezenkívül az adatok keresése és elemzése egy ilyen adattárban jelentős erőfeszítést igényel, különösen, ha heterogén információkról van szó. E korlátok leküzdésére, valamint a hagyományos adattárházak és adattavak legjobb tulajdonságainak ötvözésére fejlesztették ki a Data Lakehouse architektúrát.

Data Lakehouse architektúra: a raktárak és adattavak szinergiája

A DWH (strukturált, kezelhető, nagy teljesítményű analitika) és a Data Lake (skálázhatóság, heterogén adatok kezelése) legjobb tulajdonságainak ötvözésére fejlesztették ki a Data Lakehouse megközelítést. Ez az architektúra egyesíti az adattavak rugalmasságát a hagyományos raktárakra jellemző nagy teljesítményű feldolgozási és kezelési eszközökkel, egyensúlyt teremtve a tárolás, az analitika és a gépi tanulás között. A Data Lakehouse az adattavak és az adattárházak szintézise, amely az előbbiek rugalmasságát és skálázhatóságát az utóbbiak kezelhetőségével és lekérdezés-optimalizálásával ötvözi.

A Data Lakehouse egy olyan architektúrális megközelítés, amely az adattavak rugalmasságát és skálázhatóságát az adattárházak kezelhetőségével és lekérdezési teljesítményével kívánja ötvözni (8.1-9. ábra).

A Data Lakehouse fő jellemzői a következők:

- **Nyílt adattárolási formátum:** a nyílt adattárolási formátumok, mint például az Apache Parquet használata hatékonyságot és optimalizált lekérdezéseket biztosít.
- **Csak olvasható séma:** a DWH hagyományos, csak írható séma megközelítésével ellentétben a Lakehouse támogatja a csak olvasható sémát, amely nagyobb rugalmasságot biztosít az adatszerkezet kezelésében.
- **Rugalmas és skálázható:** támogatja a strukturált és strukturálatlan adatok tárolását és elemzését, a tárolási szintű optimalizálás révén nagy lekérdezési teljesítményt biztosít.

A Data Lakehouse olyan kompromisszumos megoldást kínál, amely egyesíti mindenki megközelítés előnyeit, így ideális a modern analitikai munkaterhelésekhez, amelyek rugalmasságot igényelnak az adatfeldolgozásban.



Ábra 8.1-9 Data Lakehouse a tárolórendszerök következő generációja, amelyet úgy terveztek, hogy megfeleljön az összetett és folyamatosan változó követelményeknek.

A modern adattárházak alapgondolata egyszerűnek tűnik: ha minden adat egy helyen van, könnyebb elemezni. A gyakorlatban azonban nem megy minden ilyen simán. Képzeljük el, hogy egy vállalat úgy dönt, hogy teljesen elhagyja a szokásos könyvelési és irányítási rendszereket (ERP, PMIS, CAFM vagy más), és egyetlen hatalmas adattóccal helyettesíti őket, amelyhez mindenki hozzáférhet. Mi fog történni? Valószínűleg káosz alakul ki: az adatok megkettőződnek, ellentmondásosak lesznek, és a kritikus információk elvesznek vagy sérülnek. Még ha az adottó csak analitikai célokra kerül is felhasználásra, megfelelő irányítás nélkül súlyosan veszélybe kerül:

- Az adatokat nehéz megérteni: a hagyományos rendszerekben az adatoknak világos struktúrája van, de egy tóban csak fájlok és táblázatok hatalmas gyűjteménye. Ahhoz, hogy valamit megtaláljon, szakembernek ki kell találnia, hogy az egyes sorok és oszlopok miért felelősek.
- Az adatok pontatlanok lehetnek: ha ugyanannak az információnak több verzióját tárolják egy helyen, nehéz előzetesen, hogy melyik verzió a legfrissebb. Ennek eredményeképpen a döntések elavult vagy hibás adatok alapján születnek.
- Az adatokat nehéz előkészíteni a munkához: az adatokat nemcsak tárolni kell, hanem kényelmes formában - jelentések, grafikonok, táblázatok formájában - is be kell mutatni. A hagyományos rendszerekben ez automatikusan történik, de az adattavakban ez további feldolgozást igényel.

Ennek eredményeképpen minden egyes adattárház-koncepciónak megvannak a maga jellemzői, feldolgozási megközelítései és üzleti alkalmazásai. A hagyományos adatbázisok a tranzakciós műveletekre összpontosítanak, az adattárházak (DWH) struktúrát biztosítanak az analitikához, az adattavak (Data Lake) nyers formában tárolják az információkat, a hibrid adattárházak (Data Lakehouse) pedig egyesítik a DWH és az Data Lake előnyeit (8.1-10. ábra).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Ábra 8.1-10 DWH, Data Lake és Data Lakehouse: az adattípusok, a felhasználási forgatókönyvek, a feldolgozási módszerek és a tárolási megközelítések fő különbségei.

A tárolási architektúra kiválasztása összetett folyamat, amely az üzleti igényektől, az információmenyiségtől és az elemzési követelményektől függ. Mindegyik megoldásnak megvannak az előnyei és hátrányai: a DWH struktúrát, a Data Lake rugalmasságát, a Lakehouse pedig a kettő közötti egyensúlyt biztosítja. A szervezetek ritkán korlátozódnak egyetlen adatarchitektúrára.

A választott architektúrától függetlenül az automatizált adatkezelési rendszerek jelentősen jobbak a manuális módszereknél. A lehető legkisebbre csökkentik az emberi hibákat, felgyorsítják az információfeldolgozást, és az üzleti folyamatok minden szakaszában biztosítják az adatok átláthatóságát és nyomon követhetőségét.

És míg a gazdaság számos területén a központosított adattárházak már iparági szabvánnyá váltak, az építőiparban a helyzet továbbra is széttagolt. Az adatok itt különböző platformokon (CDE, PMIS, ERP stb.) oszlanak meg, ami megnehezíti az egységes kép kialakítását, és olyan architektúrákat igényel, amelyek képesek ezeket a forrásokat holisztkus, analitikusan használható digitális környezetbe ötvözni.

CDE, PMIS, ERP vagy DWH és Data Lake

Néhány építőipari és mérnöki vállalat már használja a közös adatkörnyezet (CDE) koncepcióját az ISO 19650 szerint. A CDE lényegében ugyanazokat a funkciókat látja el, mint az adattárház (DWH) más iparágakban: az információk központosítása, verzióellenőrzés, validált információkhoz való hozzáférés biztosítása.

A közös adatkörnyezet (CDE) egy központosított digitális tér, amelyet a projektinformációk kezelésére, tárolására, megosztására és együttműködésére használnak a létesítmény életciklusának minden fázisában. A CDE-t gyakran felhőalapú technológiák felhasználásával valósítják meg, és integrálják a CAD (BIM) rendszerekkel.

A pénzügyi, kiskereskedelmi, logisztikai és ipari ágazatokban már évtizedek óta használnak központosított adatkezelési rendszereket, amelyek különböző forrásokból származó információkat egyesítnek, ellenőrzik azok relevanciáját és elemzési lehetőségeket biztosítanak. A CDE továbbviszi ezeket az elveket, és az épületek tervezésének és életciklus-menedzsmentjének kihívásaihoz igazítja őket.

A DWH -hoz hasonlóan a CDE is strukturálja az adatokat, rögzíti a változásokat, és egyetlen ponton keresztül biztosítja az ellenőrzött információkhoz való hozzáférést. A felhőbe való költözéssel és az analitikai eszközökkel való integrációval a kettő közötti különbségek egyre kevésbé tűnnek fel. Ha a CDE granuláris adatokkal egészül ki, amelynek koncepcióját a CAD -szolgáltatók 2023[93, 125] óta tárgyalják, még több párhuzamot láthatunk a klasszikus DWH -val.

Az "Építőipari ERP és PMIS rendszerek" című fejezetben korábban már áttekintettük a PMIS (projektmenedzsment információs rendszer) és az ERP (Enterprise Resource Planning) rendszereket. Az építési projektekben a CDE és a PMIS együtt működik: a CDE az adatok, köztük a rajzok, modellek és a projektdokumentáció tárolására szolgál, míg a PMIS olyan folyamatokat kezel, mint a határidők, feladatok, erőforrások és költségvetések ellenőrzése.

Az ERP, amely a vállalkozás egészének (pénzügy, beszerzés, személyzet, termelés) irányításáért felelős, integrálható a PMIS rendszerrel, amely vállalati szintű költség- és költségvetési ellenőrzést biztosít. Az analitika és a jelentéstétel céljából a DWH használható a CDE, a PMIS és az ERP adatainak összegyűjtésére, strukturálására és aggregálására a pénzügyi KPI-k (ROI) értékelése és a minták azonosítása érdekében. Az adattó (DL) viszont kiegészítheti a DWH-t a nyers és strukturálatlan adatok (pl. naplók, érzékelőadatok, képek) tárolásával. Ezek az adatok feldolgozhatók és betölthetők a DWH-ba további elemzés céljából.

Így a CDE és a PMIS a projektmenedzsmentre, az ERP az üzleti folyamatokra, a DWH és az Data Lake pedig az analitikára és az adatműveletekre összpontosít.

A CDE, PMIS és ERP rendszerek összehasonlítása során a DWH és Data Lake rendszerekkel jelentős különbségek mutatkoznak a szállítói függetlenség, a költségek, az integrációs rugalmasság, az adatfüggetlenség, a változásokhoz való alkalmazkodás sebessége és az elemzési képességek tekintetében (8.1-11. ábra). Az olyan hagyományos rendszerek, mint a CDE, a PMIS és az ERP gyakran kötődnek bizonyos szállítói megoldásokhoz és szabványokhoz, ami csökkenti a rugalmasságukat, és a licencek és a támogatás miatt növeli a költségeiket. Ezen túlmenően az ilyen rendszerekben lévő adatok gyakran védett, zárt formátumban vannak tárolva, ami korlátozza a felhasználást és az elemzést.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Ábra 8.1-11 DWH és Data Lake nagyobb rugalmasságot és adatfüggetlenséget kínálnak, mint az olyan rendszerek, mint a CDE, PMIS és ERP.

Ezzel szemben a DWH és a Data Lake nagyobb rugalmasságot biztosít a különböző adatforrásokkal való integrálásban, és a nyílt technológiák és platformok használata segít csökkenteni a teljes tulajdonlási költséget. A DWH és a Data Lake emellett az elemzési eszközök széles skáláját támogatja, ami növeli az elemzési és kezelési képességeket.

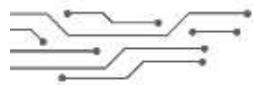
A CAD-formátumok visszafejtésére szolgáló eszközök fejlődésével és a CAD-alkalmazások adatbázisaihoz való hozzáféréssel egyre égetőbbé válik a kérdés: mennyire indokolt továbbra is zárt, elszigetelt platformokat használni, ha a tervezési adatoknak a szakemberek széles köre számára elérhetőnek kell lenniük, akik tucatnyi vállalkozó és tervező szervezetnél dolgoznak?

Ez a gyártóspecifikus technológiai függőség jelentősen korlátozhatja az adatkezelés rugalmasságát, lassíthatja a projektváltozásokra való reagálást, és gátolhatja a résztvevők közötti hatékony együttműködést.

Az adatkezelés hagyományos megközelítései - beleértve a DWH, Data Lake, CDE és PMIS - elsősorban az információk tárolására, strukturálására és feldolgozására összpontosítottak. A mesterséges intelligencia és a gépi tanulás fejlődésével azonban egyre nagyobb szükség van az adatok rendszerezésének új módozataira, amelyek nemcsak összesítik, hanem fel is tárgák az összetett összefüggéseket, megtalálják

a rejtett mintákat, és azonnali hozzáférést biztosítanak a legfontosabb információkhoz.

A vektoros adatbázisok - a nagydimenziós beágyazásokra optimalizált új típusú tárolók - kezdenek különleges szerepet játszani ebben az irányban.



FEJEZET 8.2.

ADATTÁRHÁZ-KEZELÉS ÉS KÁOSZMEGELŐZÉS

Vektoradatbázisok és a határoló doboz

A vektoradatbázisok a tárolók új osztályát jelentik, amelyek nem csupán adatokat tárolnak, hanem lehetővé teszik a jelentés szerinti keresést, az objektumok szemantikai közelsg szerinti összehasonlítását, valamint intelligens rendszerek létrehozását: az ajánlásuktól az automatikus elemzésig és a kontextusgenerálásig. A hagyományos adatbázisokkal ellentétben, amelyek a pontos egyezésekre összpontosítanak, a vektoradatbázisok az attribútumok alapján találnak hasonló objektumokat - még akkor is, ha nincs pontos egyezés

A vektoradatbázis az adatbázisok egy speciális típusa, amely az adatokat többdimenziós vektorok formájában tárolja, amelyek mindegyike bizonyos jellemzőket vagy tulajdonságokat képvisel. Ezek a vektorok az adatok összetettségétől függően különböző számú dimenzióval rendelkezhetnek (egyik esetben néhány dimenzióról, másik esetben — ezerről lehet szó).

A vektoradatbázisok fő előnye, hogy a keresés az értékek pontos egyezése helyett a szemantikai relevancia alapján történik. Az SQL- és Pandas -lekérdezések helyett az "egyenlő" vagy "tartalmazza" szűrőkkel a legközelebbi szomszédok (k-NN) keresése (a könyv következő részében bővebben beszélünk a k-NN-ről) a jellemzőtérből.

Az LLM (Large Language Models) és a generatív modellek fejlesztésével az adatbázisokkal való interakció kezd megváltozni. Lehetővé vált az adatok természetes nyelven történő lekérdezése, szemantikus keresés a dokumentumokon, a kulcskifejezések automatikus kinyerése és az objektumok közötti kontextuális kapcsolatok kialakítása - minden az SQL ismerete vagy a táblázatok szerkezetének ismerete nélkül. Ezt részletesebben az "LLM-ek és szerepük az adatfeldolgozásban és az üzleti folyamatokban" című részben tárgyalunk.

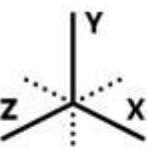
Fontos azonban tudatosítani, hogy az LLM-ek nem strukturálják és rendszerezik automatikusan az információkat. A modell csak átfut az adatokon, és a lekérdezés kontextusa alapján megtalálja a legrélevánsabb adatot. Ha az adatok nincsenek előzetesen megtisztítva vagy átalakítva, a mélykeresés olyan lesz, mintha a digitális "szemétben" próbálnánk választ találni - lehet, hogy működik, de az eredmények minősége alacsonyabb lesz. Ideális esetben, ha az adatok strukturálhatók (pl. a dokumentumok Markdownra fordítása) és betölthetők egy vektoros adatbázisba. Ez jelentősen növeli a kimenet pontosságát és relevanciáját.

A vektoradatbázisokat kezdetben a gépi tanulásban használták, de ma már egyre több alkalmazásuk van ezen kívül is - a keresőmotorokban, a tartalom személyre szabásában és az intelligens analitikában.

A vektoros megközelítés egyik legnyilvánvalóbb példája a konstrukcióban a Bounding Box (határoló paralelepiped). Ez egy geometriai konstrukció, amely egy objektum határait írja le a háromdimenziós térben. A Bounding Boxot a minimális és maximális X, Y és Z koordináták határozzák meg, egy "dobozt" alkotva az

objektum körül. Ez a módszer lehetővé teszi egy elem méretének és elhelyezésének becslését anélkül, hogy a teljes geometriát elemeznie kellene.

Minden Bounding Box egy többdimenziós térben vektorként ábrázolható: például [x, y, z, szélesség, magasság, mélység] - már 6 dimenzió (8.2-1 ábra).



Bounding Box

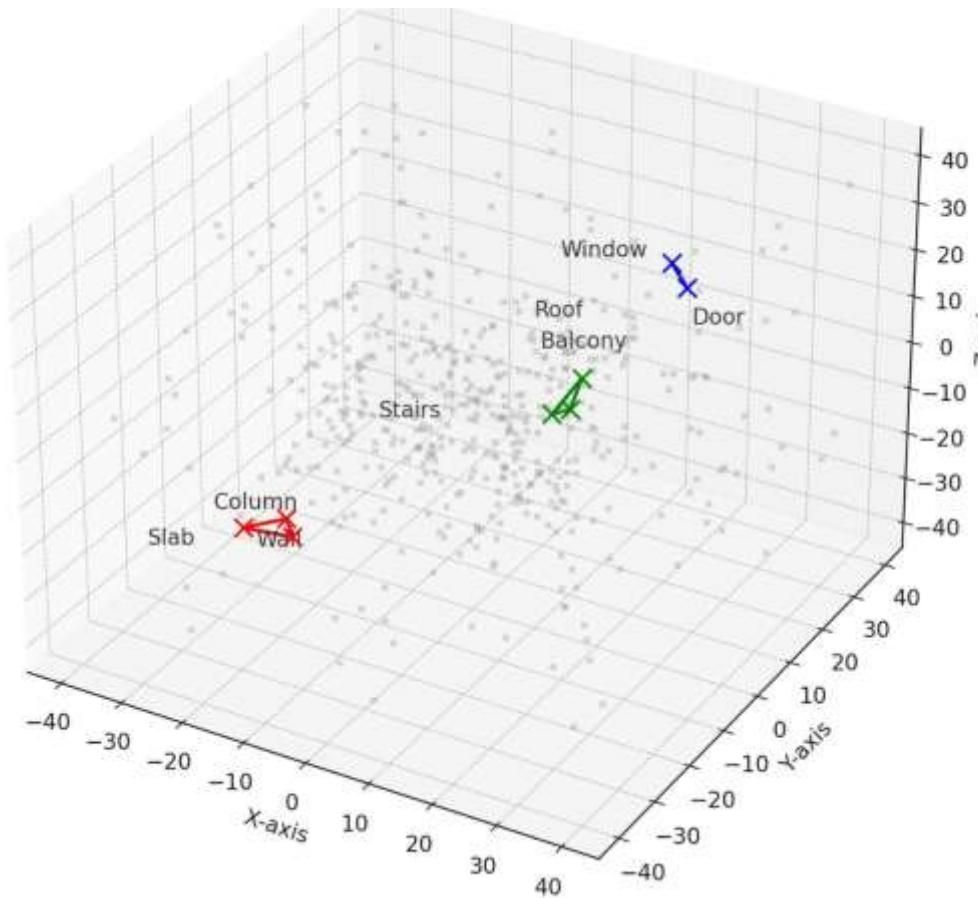
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

8.2-1 ábra Bounding Box -elem koordinátaadatai és elhelyezkedésük a projektmodellben egy vektoradatbázishoz hasonlóan.

Ez az adatreprézentáció számos feladatot megkönnyít, például az objektumok közötti metszéspontok ellenőrzését, az épületelemek térfelületi elosztásának tervezését és az automatikus számítások elvégzését. A Bounding Box hídként szolgálhat az összetett 3D modellek és a hagyományos vektoros adatbázisok között, lehetővé téve mindenekkel megközelítés előnyeinek hatékony kihasználását az építészeti és mérnöki modellezésben

A Bounding Box a geometria "vektorizálása", a beágyazás (valami absztrakt dolog átalakításának módja) pedig a "jelentés vektorizálása". Mindkét megközelítés lehetővé teszi, hogy a kézi keresésről az intelligens keresésre térjünk át, legyen szó 3D -objektumokról egy projektmodellben vagy fogalmakról egy szövegben.

Az objektumok keresése a projektben (például "keress minden ablakot, amelynek szélessége > 1,5 m") hasonló a legközelebbi szomszédok kereséséhez (k-NN) egy vektoros adatbázisban, ahol a kritériumok egy "zónát" határoznak meg a jellemzőtérből. (a k-NN legközelebbi szomszédok kereséséről a következő, gépi tanulásról szóló részben lesz szó bővebben) (8.2-2. ábra). Ha további paramétereket (anyag, súly, gyártási idő) adunk a határoló doboz attribútumaihoz, a táblázat egy magas dimenziós vektorról alakul, ahol minden attribútum egy új dimenziót jelent. Ez közelebb áll a modern vektorbázisokhoz, ahol a dimenziókat százakban vagy ezrekben számolják (pl. a neurális hálózatokból származó beágyazása).



Ábra 8.2-2 Objektumok keresése a projektben vektoros adatbázisok segítségével.

A Bounding Box, nemcsak geometriai objektumokra, hanem szöveg- és nyelvi elemzésekre is alkalmazható. Az adatok vektoros reprezentációját már aktívan használják a természetes nyelvi feldolgozásban (NLP). Ahogyan egy építési projektben az objektumok csoportosíthatók térbeli közelsgük alapján (8.2-2. ábra), úgy a szövegben lévő szavak is elemezhetők szemantikai és kontextuális közelsgük alapján.

Például az "építész", "építés", "tervezés" szavak egymás mellett lesznek a vektortérben, mert hasonló jelentéssel bírnak. Az LLM-ben ez a mechanizmus automatikus, kézi kategorizálást nem igénylő kategorizálást tesz lehetővé:

- A szöveg téma-jának azonosítása
- Szemantikus keresés a dokumentumok tartalmában
- Automatikus megjegyzések és szöveges összefoglalók készítése
- Szinonimák és kapcsolódó kifejezések keresése

A vektoros adatbázisok lehetővé teszik a szöveg elemzését és a kapcsolódó kifejezések megtalálását, ugyanúgy, ahogyan a Bounding Box segít a térbeli objektumok elemzésében a 3D-modellekben. A projektelemek Bounding Box példája segít megérteni, hogy a vektoros ábrázolás nem egy tisztán "mesterséges" fogalom az ML-ből, hanem az adatok strukturálásának természetes módja az alkalmazott problémák megoldására, legyen szó akár oszlopok kereséséről egy CAD-projektben, akár szemantikailag

közéllálló képekről egy adatbázisban.

Az adatbázisokkal dolgozó szakembereknek figyelmet kell fordítaniuk a vektortárolókra. Elterjedésük az adatbázisok fejlődésének új szakaszát jelzi, ahol a klasszikus relációs rendszerek és az AI-orientált technológiák kezdenek összefonódni, és a jövő hibrid megoldásait alkotják.

Az összetett és nagyméretű mesterséges intelligencia -alkalmazásokat fejlesztő felhasználók speciális adatbázisokat használnak a vektoros kereséshez. Ugyanakkor azok, akiknek csak külön AI-funkciókra van szükségük a meglévő alkalmazásokba való integráláshoz, nagyobb valószínűséggel választják a már használt adatbázisok (PostgreSQL, Redis) beépített vektoros keresési képességeit.

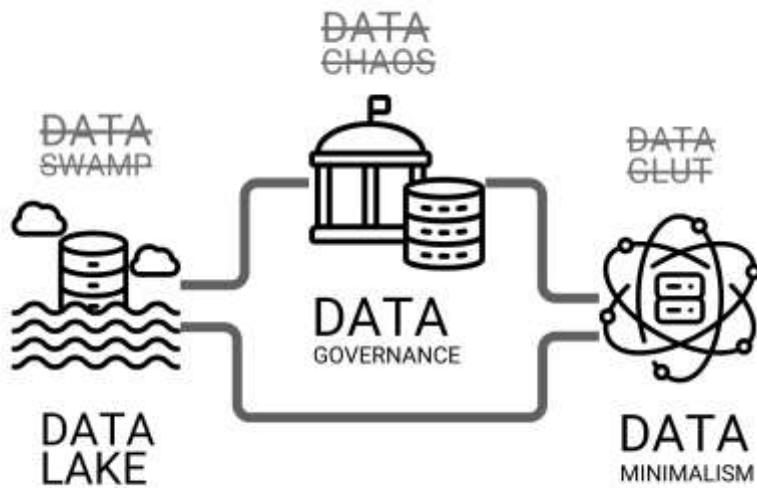
Bár az olyan rendszerek, mint a DWH, Data Lake, CDE, PMIS, vektoradatbázisok és mások különböző megközelítéseket kínálnak az adatok tárolására és kezelésére, hatékonyságukat nemcsak az architektúrájuk határozza meg, hanem az is, hogy maga az adat mennyire jól van megszervezve és kezelve. Még a modern megoldások használata esetén is - legyenek azok vektoros adatbázisok, klasszikus relációs DBMS-ek vagy Data Lake típusú raktárak - az adatok kezelésére, strukturálására és frissítésére vonatkozó egyértelmű szabályok hiánya ugyanazokhoz a nehézségekhez vezethet, amelyekkel a különböző fájlokkal és több formátumú adatokkal dolgozó felhasználóknak is szembe kell nézniük.

Adatkezelés nélkül), még a leghatékonyabb megoldások is kaotikussá és strukturálatlanná válhatnak, és az adattavak adatmocsárrá válhatnak). Ennek elkerülése érdekében a vállalatoknak nemcsak a megfelelő tárolási architektúrát kell kiválasztaniuk, hanem adatminimalizmust), hozzáférés-kezelési és minőségellenőrzési stratégiákat is be kell vezetniük, hogy az adatokat hatékony döntéshozatali eszközökkel alakítsák.

(Adatkezelés), adatminimalizmus) és adatmocsár)

Az adatkezelés), az adatminimalizmus) és az adatmocsár megelőzése) fogalmak megértése és végrehajtása kulcsfontosságú az adattárházak sikeres kezeléséhez és az üzleti érték előállításához (8.2-3. ábra).

A Gartner (2017) tanulmánya szerint a nagy adatprojektek 85%-a kudarcot vall, és ennek egyik fő oka a nem megfelelő adatminőség és adatkezelés [144].



Ábra 8.2-3 Az adatkezelés néhány kulcsfontosságú szempontja az adatkezelés és az adatminimalizmus.

Az adatkezelés (Data Governance) az adatkezelés alapvető eleme, amely biztosítja, hogy az adatokat megfelelően és hatékonyan használják fel az összes üzleti folyamatban. Nemcsak a szabályok és eljárások kialakításáról szól, hanem az adatok rendelkezésre állásának, megbízhatóságának és biztonságának biztosításáról is:

- A adatok meghatározása és osztályozása: Az entitások egyértelmű meghatározása és osztályozása lehetővé teszi a szervezetek számára, hogy megértsék, milyen entitásokra van szükség a vállalatnál, és meghatározzák, hogyan kell használni őket.
- Hozzáférési jogok és kezelés: az adatokhoz való hozzáférésre és kezelésre vonatkozó irányelvek és eljárások kidolgozása biztosítja, hogy csak az arra jogosult felhasználók férhessenek hozzá bizonyos adatokhoz.
- Az adatok védelme a külső fenyegetésekkel szemben: Az adatok védelme a külső fenyegetésekkel szemben az adatkezelés kulcsfontosságú szempontja. Ez nemcsak technikai intézkedéseket foglal magában, hanem az alkalmazottak képzését is az információbiztonság alapjairól.

Az adatminimalizmus (Data Minimalism) olyan megközelítés, amely az adatokat a legértékesebb és legjelentősebb attribútumokra és entitásokra csökkenti a formációban (8.2-4. ábra), ezáltal csökkenti a költségeket és javítja az adatok felhasználását:

- A döntéshozatal egyszerűsítése: az objektumok és attribútumaik számának csökkentése a legfontosabbakra egyszerűsíti a döntéshozatalt azáltal, hogy csökkenti az adatok elemzéséhez és feldolgozásához szükséges időt és erőforrásokat.
- Koncentrálás a fontos dolgokra: a legrélevánsabb entitások és attribútumok kiválasztása lehetővé teszi, hogy az üzlet szempontjából valóban fontos információkra összpontosítson, kiküszöböльve a zajt és a felesleges adatokat.
- Hatékony erőforrás-elosztás: az adatok minimalizálása hatékonyabb erőforrás-elosztást tesz lehetővé, csökkenti az adattárolási és feldolgozási költségeket, javítja az adatok minőségét és biztonságát.

Az adatokkal való munka logikájának nem az adatok létrehozásával kell kezdődni (8.2-4. ábra), hanem az

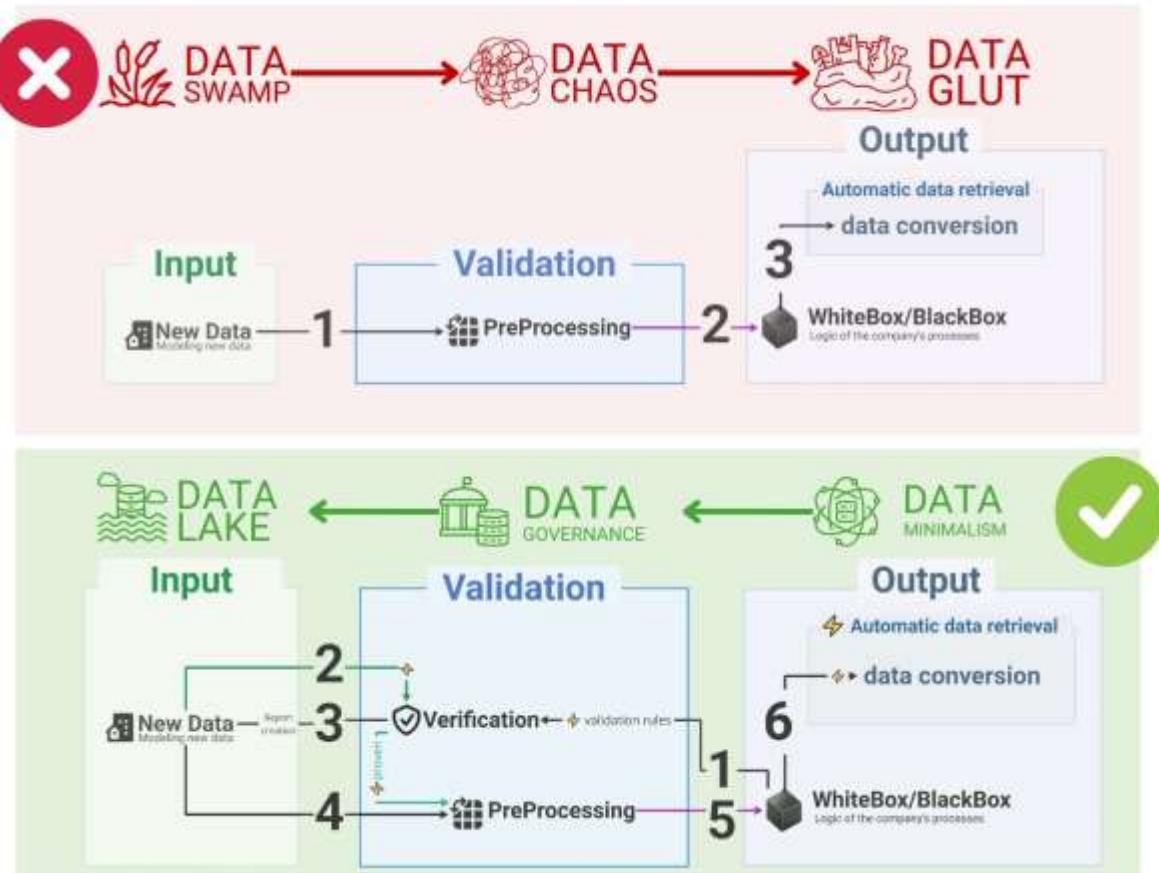
adatok felhasználásának jövőbeli forgatókönyveinek megértésével, még a generálási folyamat megkezdése előtt. Ez a megközelítés lehetővé teszi az attribútumokra, azok típusaira és határértékeire vonatkozó minimálisan szükséges követelmények előzetes meghatározását. Ezek a követelmények képezik az alapját a helyes és stabil entitások létrehozásának az információs modellben. Az adatok céljainak és felhasználásának előzetes megértése hozzájárul az elemzésre alkalmas struktúra kialakításához. Az adatmodellezés fogalmi, logikai és fizikai szintű megközelítéseivel kapcsolatos további részleteket az "Adatmodellezés: fogalmi, logikai és fizikai modell" című fejezetben tárgyaltuk.

Az építőipari vállalatok hagyományos üzleti folyamataiban az adatfeldolgozás gyakrabban hasonlít az adatok mocsárba dobására, ahol az adatokat először létrehozzák, majd a szakemberek megpróbálják integrálni azokat más rendszerekbe és eszközökbe.

Az adatsár (Data Swamp) az adatok ellenőrizetlen gyűjtésének és tárolásának eredménye, megfelelő szervezés, strukturálás és kezelés nélkül, ami strukturálatlan, nehezen felhasználható és kevéssé értékes adatokat eredményez.

Hogyan lehet megakadályozni, hogy az információáramlás mocsárrá váljon:

- **Adatszerkezet-kezelés:** a adatok strukturálásának és kategorizálásának biztosítása segít megelőzni az adatok elárasztását azáltal, hogy rendezetté és könnyen hozzáférhetővé teszi azokat.
- Az adatok **megértése és értelmezése:** az adatok eredetének, módosításainak és jelentésének világos leírása biztosítja az adatok megértését és helyes értelmezését.
- **Az adatminőség fenntartása:** az adatok rendszeres karbantartása és tisztítása segít fenntartani az adatok minőségét, relevanciáját és értékét az elemzési és üzleti folyamatok számára.



8.2-4. ábra Az adattárház zsúfoltságának elkerülése érdekében az adatok létrehozásának folyamatát az attribútumkövetelmények összegyűjtésével kell kezdeni.

Az adatkezelés és az adatminimalizmus elveinek az adatkezelési folyamatokba való integrálásával, valamint annak aktív megakadályozásával, hogy az adattárházak adatmocsárrá váljanak, a szervezetek maximalizálhatják az adataikban rejlő lehetőségeket.

Az adatkezelés és a minimalizmus kérdésének megoldása után az adatokkal való munka fejlődésének következő szakasza az automatikus feldolgozás szabványosítása, a minőségbiztosítás és olyan módszerek bevezetése, amelyek az adatokat elemzés, átalakítás és döntéshozatal számára használhatóvá teszik. Erre szolgálnak a DataOps és a VectorOps módszertanok, amelyek a nagy adatokkal és a gépi tanulással dolgozó vállalatok számára fontos eszközökévé válnak.

DataOps és VectorOps: új adatszabványok

Míg az adatkezelés az adatok ellenőrzéséért és rendszerezéséért felelős, addig az adatirányítás segít biztosítani azok pontosságát, konzisztenciáját és zökkenőmentes áramlását a vállalaton belül. Ez különösen kritikus az építőipar számos üzleti esete esetében, ahol az adatok folyamatosan keletkeznek és időben történő feldolgozást igényelnek. Például olyan helyzetekben, amikor az épületinformációs modellek, a projektkövetelményeket és az analitikai jelentéseket egyetlen munkanapon belül

szinkronizálni kell a különböző rendszerek között, a DataOps szerepe kulcsfontosságú lehet. Lehetővé teszi stabil és megismételhető adatfeldolgozási folyamatok kialakítását, csökkentve a késedelmek és az információk relevanciájának elvesztésének kockázatát.

Az adatkezelés önmagában nem elég - elengedhetetlen, hogy az adatokat ne csak tárolják, hanem aktívan használják is a napi működés során. Itt jön a képbe a DataOps - az automatizálásra, integrációra és folyamatos adatáramlásra összpontosító módszertan.

A DataOps az adatáramlás együttműködésének, integrációjának és automatizálásának javítására összpontosít a szervezetekben. A DataOps gyakorlatok alkalmazása elősegíti az adatok pontosságát, konzisztenciáját és rendelkezésre állását, ami kritikus fontosságú az adatközpontú alkalmazások számára.

A DataOps ökoszisztéma kulcsfontosságú eszközei az Apache Airflow (7.4-4. ábra) a munkafolyamatok összehangolásához, valamint az Apache NiFi (7.4-5. ábra) az adatfolyamok útválasztásához és átalakításához. Ezek a technológiák együtt rugalmas, megbízható és skálázható adatvezetéket tesznek lehetővé a rendszerek közötti automatikus feldolgozáshoz, ellenőrzéshez és az információk integrálásához (további részletek az "Automatikus ETL -szállító " fejezetben). A DataOps megközelítés építési folyamatokban történő megvalósításakor négy alapvető szempontot kell figyelembe venni:

- Az emberek és az eszközök fontosabbak, mint az adatok:** a silózott adattárak nagy problémának tűnhet, de a valóság ennél összetettebb. Az adatok széttagoltsága mellett a csapatok elszigeteltsége és az általuk használt eltérő eszközök is jelentős szerepet játszanak. Az építőiparban különböző szakterületek szakemberei dolgoznak az adatokkal: adatmérnökök és elemzők, BI- és vizualizációs csapatok, valamint projektmenedzsment- és minőségügyi szakértők. Mindegyiküknek más-más a munkamódszere, ezért fontossá válik egy olyan ökoszisztéma létrehozása, ahol az adatok szabadon áramlanak a résztvevők között, és az információk egységes, konzisztens változatát biztosítják.
- Automatizálja a tesztelést és a hibák felderítését:** Az építési adatok minden tartalmaznak hibákat, legyen szó a modellek pontatlanságáról, számítási hibákról vagy elavult specifikációkról. Az adatok rendszeres tesztelése és a visszatérő hibák kiküszöbölése jelentősen javíthatja az adatminőséget. A DataOps részeként olyan automatizált ellenőrzéseket és érvényesítési mechanizmusokat kell bevezetnie, amelyek figyelemmel kísérik az adatok helyességét, elemzik a hibákat és azonosítják a mintákat, valamint minden munkafolyamatban rögzítik és kezelik a rendszerhibákat. Minél magasabb az automatizált érvényesítés mértéke, annál jobb az általános adatminőség, és annál kisebb a hibák valószínűsége a végső fázisokban.
- Az adatokat ugyanúgy kell tesztelni, mint a programkódot:** a legtöbb építési alkalmazás adatfeldolgozáson alapul, de annak ellenőrzése gyakran másodlagos szerepkörökre marad. Ha a gépi tanulási modelleket pontatlan adatokon képzik ki, az hibás előrejelzésekhez és pénzügyi veszteségekhez vezet. A DataOps keretében az adatokat ugyanolyan szigorú ellenőrzésnek kell alávetni, mint a szoftverkódot: logikai ellenőrzések, stressztesztek, a modellek viselkedésének értékelése, amikor a bemeneti értékek megváltoznak. Csak hitelesített és megbízható adatokat lehet a vezetői döntések alapjául használni.
- Az adatok megfigyelhetősége a teljesítmény veszélyeztetése nélkül:** az adatfigyelés nem csupán mérőszámok gyűjteménye, hanem stratégiai minőségirányítási eszköz. Ahhoz, hogy a DataOps hatékonyan működjön, a megfigyelhetőséget az adatkezelés minden szakaszába be kell építeni, a tervezéstől az üzemeltetésig. Ugyanakkor fontos, hogy a megfigyelés ne lassítsa le a rendszert. Az

építési projektek nél kritikus fontosságú, hogy ne csak adatokat gyűjtsünk, hanem ezt úgy tegyük, hogy az adatokat létrehozó szakemberek (pl. tervezők) munkáját semmilyen módon ne zavarjuk meg. Ez az egyensúly lehetővé teszi az adatminőség ellenőrzését a termelékenység veszélyeztetése nélkül.

A DataOps nem egy további teher az adattudósok számára, hanem munkájuk gerince. A DataOps bevezetésével az építőipari vállalatok a kaotikus adatkezelésből egy olyan hatékony ökoszisztémába léphetnek át, ahol az adatok az üzletért dolgoznak.

A VectorOps viszont a DataOps fejlődésének következő lépcsőfokát jelenti, amely a többdimenziós vektoros adatok feldolgozására, tárolására és elemzésére összpontosít (amelyeket az előző fejezetben tárgyaltunk). Ez különösen fontos az olyan területeken, mint a digitális ikrek, a neurális hálózati modellek és a szemantikus keresés, amelyek kezdenek eljutni az építőiparba. A VectorOps a vektoradatbázisokra támaszkodik az objektumok többdimenziós reprezentációinak hatékony tárolása, indexelése és keresése érdekében.

A VectorOps a DataOps után a következő lépés, amely a vektoros adatok feldolgozására, elemzésére és felhasználására összpontosít az építőiparban. A DataOps-szal ellentétben, amely az adatáramlásra, a konzisztenciára és a minőségre összpontosít, a VectorOps a gépi tanuláshoz szükséges többdimenziós objektumreprezentációk kezelésére összpontosít.

A hagyományos megközelítésekkel eltérően a VectorOps lehetővé teszi a pontosabb objektumleírások elérését, ami kritikus fontosságú a digitális ikrek, a generatív tervezési rendszerek és a vektoros formátumba konvertált CAD-adatok automatikus hibafelismerése szempontjából. A DataOps és a VectorOps együttes megvalósítása szilárd alapot képez a nagy mennyiséggű információval végzett skálázható, automatizált munkához - a klasszikus táblázatoktól a szemantikailag gazdag térbeli modellekig.

A következő lépések: a kaotikus tárolástól a strukturált tárolásig

Az adattárolás hagyományos megközelítései gyakran vezetnek olyan széttagolt "információs silók" létrehozásához, ahol a fontos információk nem hozzáférhetők az elemzéshez és a döntéshozatalhoz. A modern tárolási koncepciók, mint például az adattárház, az adattó és ezek hibridjei lehetővé teszik a széttagolt információk egységesítését és központosított módon történő rendelkezésre bocsátását az adatfolyam és az üzleti intelligencia számára. Nemcsak a megfelelő tárolási architektúra kiválasztása fontos, hanem az adatkezelés (és az adatminimalizmus) megvalósítása is, hogy a tároló létesítmények ne váljanak ellenőrizhetetlen adatmocsárrá).

E rész összegzéseként érdemes kiemelni a főbb gyakorlati lépéseket, amelyek segítenek a tárgyalt koncepciókat a napi feladatokban alkalmazni:

- Hatékony adattárolási formátumok kiválasztása
 - A CSV és az XLSX helyett hatékonyabb formátumok (Apache Parquet, ORC) használata nagy mennyiséggű adat tárolására.
 - Adatverzió-nyilvántartó rendszer bevezetése a változások nyomon követésére
 - Metaadatok használata az információ szerkezetének és eredetének leírására

■ Egységes vállalati adatarchitektúra létrehozása

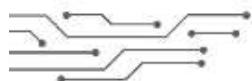
- Különböző tárolási architektúrák összehasonlítása: RDBMS, DWH és Data Lake. Válassza ki azt, amelyik a legjobban megfelel a skálázhatósági, forrásintegrációs és analitikai feldolgozási igényeinek.
- Tervezzen folyamatábrát az adatok különböző forrásokból történő kinyerésére, betöltésére és átalakítására (ETL) a feladataihoz. Használjon olyan vizualizációs eszközöket, mint a Miro, a Lucidchart vagy a Draw.io a kulcsfontosságú lépések és integrációs pontok megjelenítéséhez.

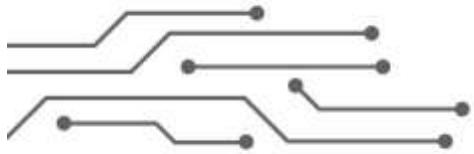
■ Adatkezelési gyakorlatok bevezetése

- Kövesse az adatminimalizmus megközelítést - csak azt tárolja és dolgozza fel, ami valóban értékes.
- Az adatkezelési elvek végrehajtása - az adatokért való felelősség meghatározása, a minőség és az átláthatóság biztosítása.
- Tudjon meg többet az adatkezelési irányelvekről és a DataOps koncepciókról, VectorOps
- Az adatminőségi kritériumok meghatározása és az adatok DataOps-on belüli validálására vonatkozó eljárások meghatározása.

A jól szervezett adattárolás megteremti az alapot a vállalat elemzési folyamatainak központosításához. A fájlok kaotikus felhalmozásáról a strukturált tárolásra való áttérés lehetővé teszi, hogy az információkból stratégiai értéket hozzanak létre, amely segít a megalapozott döntések meghozatalában és az üzleti folyamatok hatékonyságának javításában.

Miután az adatgyűjtési, -átalakítási, -elemzési és strukturált tárolási folyamatokat automatizálták és szabványosították, a digitális átalakulás következő szakasza a Big Data teljes körű kezelése.





IX RÉSZ

BIG DATA, GÉPI TANULÁS ÉS ELŐREJELZÉSEK

A kilencedik rész a nagy adatokra, a gépi tanulásra és a prediktív analitikára összpontosít az építőiparban. Az intuitív döntéshozatalról a múltbeli adatokon alapuló objektív elemzésekre való áttérést vizsgálja. Gyakorlati példákon keresztül mutatja be a nagy adatelemzést az építőiparban - a San Franciscó-i építési engedélyek adatállományának elemzéstől a CAD - több millió elemet tartalmazó projektek feldolgozásáig. Különös figyelmet fordítanak a gépi tanulási módszerekre az építési projektek költségeinek és ütemezésének előrejelzésére, a lineáris regresszió és a k-nearest neighbours algoritmusok részletes tárgyalása mellett. Bemutatjuk, hogy a strukturált adatok hogyan válnak a kockázatok értékelését, az erőforrások optimalizálását és a projektmenedzsment hatékonyságának javítását szolgáló előrejelző modellek alapjává. A rész ajánlásokat is ad a reprezentatív adatminták kiválasztására, és elmagyarázza, hogy a hatékony elemzésekhez miért nem mindenkor van szükség nagy adathalmazokra.

FEJEZET 9.1.

BIG DATA ÉS ANNAK ELEMZÉSE

Big data az építőiparban: az intuíciótól a kiszámíthatóságig

A "big data" kifejezésnek nincs szigorú meghatározása. A fogalom eredetileg akkor jelent meg, amikor az információ mennyisége kezdte meghaladni a feldolgozás hagyományos módszereinek képességeit. Napjainkra az adatok mennyisége és összetettsége számos iparágban, köztük az építőiparban is, annyira megnőtt, hogy nem férnek el a számítógépek helyi memóriájában, és feldolgozásukhoz új technológiák alkalmazására van szükség.

A Big Data-val való munka lényege nemcsak a tárolás és a feldolgozás, hanem a prediktív képességek is. Az építőiparban a Big Data megnyitja az utat a táblázatok és vizualizációk szubjektív értelmezésén alapuló intuitív döntésektől (amint azt korábban tárgyaltuk) a valós megfigyelésekkel és statisztikákkal alátámasztott, megalapozott előrejelzések felé.

A közhiedelemmel ellentétben a nagy adatmennyiséggel való munka célja nem az, hogy "egy gép úgy gondolkodjon, mint egy ember", hanem az, hogy matematikai modelleket és algoritmusokat alkalmazzunk hatalmas adatmennyiségek elemzésére a minták azonosítása, az események előrejelzése és a folyamatok optimalizálása érdekében.

A Big Data nem az algoritmusok hideg, emberi befolyástól mentes világa. Éppen ellenkezőleg, a nagy adatok az ösztöneinkkel, a hibáinkkal és a kreativitásunkkal együtt működnek. Az emberi gondolkodás tökéletlensége az, ami lehetővé teszi számunkra, hogy nem szokványos megoldásokat találunk és áttörést érjünk el.

A digitális technológia fejlődésével az építőipar elkezdte aktívan használni az informatikai ágazatból származó adatfeldolgozási technikákat. Az olyan eszközöknek köszönhetően, mint a Pandas és az Apache Parquet, a strukturált és strukturálatlan adatok kombinálhatók, egyszerűsítve az információkhoz való hozzáférést és csökkentve az elemzéssel járó veszteségeket, míg a dokumentumokból vagy CAD-projektekből származó nagy adathalmazok (9.2-10. ábra - 9.2-12. ábra) lehetővé teszik az adatok gyűjtését, elemzését és előrejelzését a projekt életciklusának minden szakaszában.

A Big Data átalakító hatással van az építőiparra, és többféleképpen is befolyásolhatja azt. A Big Data technológiák alkalmazása számos kulcsfontosságú területen hoz eredményeket, többek között például a következőkben:

- **Beruházási potenciálelemzés** - a projektek jövedelmezőségének és megtérülési idejének előrejelzése a korábbi létesítmények adatai alapján.
- **Előrejelző karbantartás** - a berendezések valószínűsíthető meghibásodásainak azonosítása, mielőtt azok ténylegesen bekövetkeznének, ami csökkenti az állásidőt.
- **Ellátási lánc optimalizálása** - a zavarok előrejelzése és a logisztikai hatékonyság javítása.
- **Energiahatékonysági elemzések** - az alacsony energiafelhasználású épületek tervezésének segítése.

- **Biztonsági felügyelet** - érzékelők és viselhető eszközök használata a helyszíni körülmények ellenőrzésére.
- **Minőségellenőrzés** - a folyamatszabványoknak való megfelelés valós idejű nyomon követése.
- **Munkaerő-gazdálkodás** - teljesítményelemzés és a személyzeti igények előrejelzése.

Nehéz olyan területet találni az építőiparban, ahol az adatelemzésre és az előrejelzésekre ne lenne igény. Az előrejelző algoritmusok legfőbb előnye, hogy képesek az adatok felhalmozódásával önállóan tanulni és folyamatosan javulni.

A közeljövőben a mesterséges intelligencia nem csupán az építkezőket fogja segíteni, hanem kulcsfontosságú döntésekkel fog hozni - a tervezési folyamatoktól kezdve az épület üzemeltetési kérdésekig.

Az előrejelzések generálásának és a tanulási modellek használatának módjáról a könyv következő részében, a "Gépi tanulás és előrejelzések" című részben lesz szó.

A nagy adatokkal való teljes körű munkára való áttérés az analitika megközelítésének megváltoztatását igényli. Míg az eddig vizsgált klasszikus rendszerek az ok-okozati összefüggésekre összpontosítottak, a nagy adatelemzés a statisztikai minták és összefüggések keresésére helyezi át a hangsúlyt, amelyek lehetővé teszik a rejtett kapcsolatok azonosítását és az objektumok viselkedésének előrejelzését még az összes tényező teljes megértése nélkül is.

A nagyméretű adatok megvalósíthatóságának megkérdőjelezése: korreláció, statisztika és adatmintavételezés

Hagyományosan az építés szubjektív hipotéziseken és személyes tapasztalatokon alapult. A mérnökök - bizonysos valószínűséggel - feltételezték, hogy az anyag hogyan fog viselkedni, milyen terheléseknek fog ellenállni a szerkezet, és mennyi ideig fog tartani a projekt. Ezeket a feltételezéseket a gyakorlatban tesztelték, gyakran idő, erőforrások és jövőbeli kockázatok árán.

A big data megjelenésével a megközelítés drámaian megváltozik: a döntések már nem intuitív megérzések alapján, hanem nagyméretű adathalmazok elemzésének eredményeként születnek. Az építőipar fokozatosan megszűnik az intuíció művészete lenni, és az előrejelzés pontos tudományává válik.

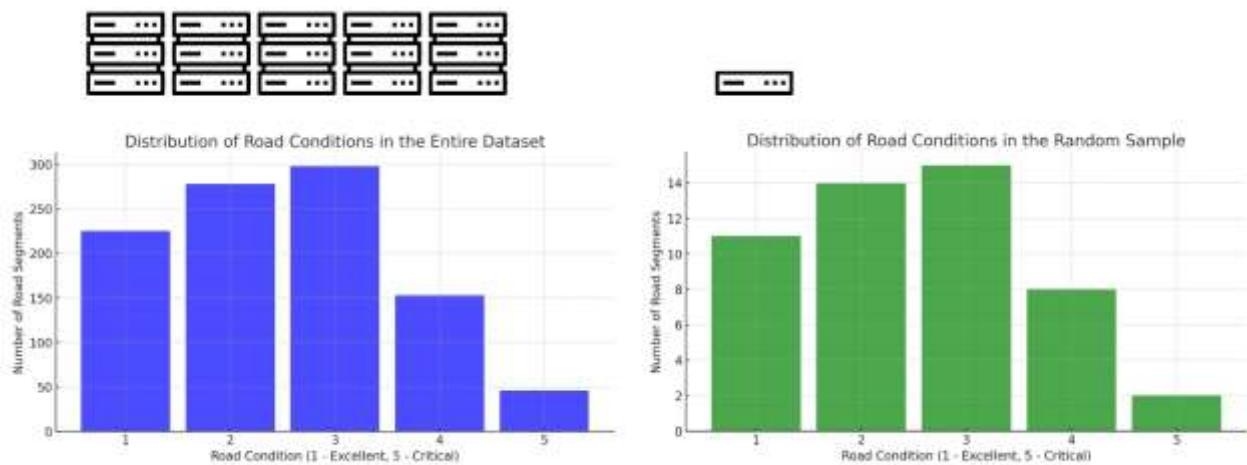
A big data használatára való áttérés elkerülhetetlenül felvet egy fontos kérdést: mennyire kritikus az adatmennyiség, és mennyi információra van valóban szükség a megbízható előrejelző elemzéshez? A széles körben elterjedt hiedelem, miszerint "minél több adat, annál nagyobb a pontosság", a gyakorlatban nem minden bizonyul statisztikailag igaznak.

Jerzy Neumann statisztikus már 1934-ben bebizonyította [145], hogy a statisztikai következtetés pontosságának kulcsa nem annyira az adatok mennyiségében, mint inkább azok reprezentativitásában és a mintavétel véletlenszerűségében rejlik.

Ez különösen igaz az építőiparban, ahol nagy mennyiségű adatot gyűjtenek IoT -érzékelők, szkennerek,

megfigyelő kamerák, drónok és még több formátumú CAD -modellek segítségével, ami növeli a vakfoltok, a kiugró értékek és az adatok torzításának kockázatát.

Vegyük egy példát az útburkolat állapotának megfigyelésére. Az összes útszakasz teljes adatkészlete X GB lehet, és körülbelül egy napot vesz igénybe a feldolgozása. Ugyanakkor egy véletlenszerű minta, amely csak minden 50. útszakaszt tartalmazna, csak $X/50$ GB-ot venne igénybe, és fél óra alatt feldolgozható lenne, miközben bizonyos számítások esetében hasonló pontosságú becsléseket adna (9.1-1. ábra).



9.1-1. ábra A járdák állapotának hisztorogramjai: a teljes adatsor és a véletlenszerű mintavétel azonos eredményeket mutat.

Így a sikeres adatelemzés kulcsa gyakran nem az adatok mennyisége, hanem a minta reprezentativitása és az alkalmazott feldolgozási módszerek minősége lehet. A véletlenszerű mintavételre és a szelektívbb megközelítésre való áttérés gondolkodásbeli változást igényel az építőiparban. Történelmileg a vállalatok a "minél több adat, annál jobb" logikát követték, abban a hitben, hogy az összes lehetséges mutató lefedése maximalizálja a pontosságot.

Ez a megközelítés emlékeztet a projektmenedzsment egy népszerű tévhítére: "minél több szakembert vonzok, annál hatékonyabb lesz a munka". A humán erőforráshoz hasonlóan azonban a minőség és az eszközök fontosabbak, mint a mennyiség. Az adatok vagy a projekt résztvevői közötti összefüggések (korrelációk) figyelembevétele nélkül a mennyiség növelése csak zajhoz, torzításhoz, duplikációhoz és felesleges pazarláshoz vezethet.

Végül gyakran kiderül, hogy sokkal produktívabb egy kisebb, de minőségileg felkészült, stabil és ésszerű előrejelzések készítésére alkalmas adathalmaz, mint a sok ellentmondásos jelet tartalmazó, masszív, de kaotikus információkra támaszkodni.

A túlzott adatmennyiség nem csak nem garantálja a nagyobb pontosságot, de torz következtetésekhez is vezethet a zaj, a redundáns jellemzők, a rejtett korrelációk és az irreleváns információk jelenléte miatt. Ilyen körülmények között megnő a modellek túlillesztésének kockázata, és csökken az analitikus eredmények megbízhatósága.

Az építőiparban a nagy mennyiségű adat kezelésének egyik legnagyobb kihívása az adatok optimális mennyiségeinek és minőségeinek meghatározása. Például a betonszerkezetek állapotának megfigyelése

során a több ezer érzékelő használata és a percenkénti információgyűjtés túlterhelheti a tároló- és elemzőrendszeret. Ha azonban korrelációs elemzést végez, és kiválasztja a 10% leginformatívabb érzékelőt, akkor az előrejelzések közel azonos pontosságúak lehetnek, sokszor, néha több tízszer és százszor kevesebb erőforrást költve.

Az adatok kisebb részhalmazának használata csökkenti mind a szükséges tárolási mennyiséget, mind a feldolgozási időt, ami jelentősen csökkenti az adatok tárolásának és elemzésének költségeit, és gyakran teszi a véletlenszerű mintavételt ideális megoldássá a prediktív elemzéshez, különösen nagy infrastrukturális projektek vagy valós idejű munka esetén. Végső soron az építési folyamatok hatékonyságát nem az összegyűjtött adatok mennyisége, hanem azok elemzésének minősége határozza meg. Kritikus megközelítés és gondos elemzés nélkül az adatok helytelen következtetésekhez vezethetnek.

Egy bizonyos adatmennyiség után minden egyes új információegység egyre kevesebb hasznos eredményt hoz. A végtelen információgyűjtés helyett fontos, hogy az információk reprezentativitására és elemzési módszereire összpontosítsunk (9.2-2. ábra).

Ezt a jelenséget jól leírja Allen Wallis [146], aki a statisztikai módszerek alkalmazását az amerikai haditengerészet két alternatív lövedéktervének tesztelésének példáján szemlélteti.

A haditengerészet két alternatív lövedékkialakítást (A és B) tesztelt párosított lövessorozattal. minden egyes menetben az A 1 vagy 0 értéket kapott, attól függően, hogy teljesítménye jobb vagy rosszabb volt-e, mint a B-é, és fordítva. A szokásos statisztikai megközelítés szerint meghatározott számú próbát végeznek (pl. 1000), és a győztest százalékos eloszlás alapján határozzák meg (pl. ha A az esetek több mint 53%-ában 1-est kap, akkor azt tekintik a legjobbnak). Amikor Allen Wallis egy ilyen problémát megvitatott Garrett L. Schuyler (haditengerészeti) kapitánnyal, a kapitány kifogásolta, hogy egy ilyen teszt - Allen beszámolóját idézve - haszontalan lehet. Ha egy olyan bölcs és tapasztalt tüzérsgé tiszt, mint Schuyler, a helyszínen lett volna, az első néhány száz [lövés] után belátta volna, hogy a kísérletet nem kell befejezni, akár azért, mert az új módszer egyértelműen gyengébb, akár azért, mert egyértelműen jobb, mint amit reméltek [146].

- Az amerikai kormány statisztikai kutatócsoportja a Columbia Egyetemen, második világháborús időszak.

Ezt az elvet széles körben alkalmazzák különböző iparágakban. Az orvostudományban például az új gyógyszerek klinikai vizsgálatát a betegek véletlenszerű mintáin végezik, ami lehetővé teszi statisztikailag szignifikáns eredmények elérését anélkül, hogy a gyógyszert a bolygón élő emberek teljes populációján kellene tesztelni. A közgazdaságtanban és a szociológiában reprezentatív felméréseket végeznek, hogy tükrözzenek a társadalom véleményét anélkül, hogy az országban mindenkit meg kellene kérdezni.

Ahogyan a kormányok és a kutatásvezetékek kis populációkat vizsgálnak az általános társadalmi tendenciák megértése érdekében, az építőipari vállalatok véletlenszerű adatmintákat használhatnak a projektmenedzsment hatékony nyomon követésére és előrejelzések készítésére (9.1-1. ábra).

A nagyméretű adatok megváltoztathatják a társadalomtudományok megközelítését, de nem fogják helyettesíteni a statisztikai józan észt [147].

- Thomas Landsall-Welfair, "A nemzet aktuális hangulatának előrejelzése", Significance v. 9(4), 2012 .

Erőforrás-takarékossági szempontból a jövőbeli előrejelzésekhez és döntéshozatalhoz szükséges adatok gyűjtésekor fontos megválaszolni a kérdést: van-e értelme jelentős erőforrásokat fordítani hatalmas adathalmazok gyűjtésére és feldolgozására, amikor sokkal kisebb és olcsóbb, fokozatosan növelhető tesztadathalmaz használható? A véletlenszerű mintavétel hatékonysága azt mutatja, hogy a vállalatok több tízszeresére vagy akár ezerszeresére csökkenthetik a modellek gyűjtésének és betanításának költségeit, ha olyan adatgyűjtési módszereket választanak, amelyek nem igényelnek átfogó lefedettséget, de mégis megfelelő pontosságot és reprezentativitást biztosítanak. Ez a megközelítés lehetővé teszi, hogy még a kisvállalatok is a nagyvállalatokkal egyenértékű eredményeket érjenek el, lényegesen kevesebb erőforrás és adatmennyiséggel felhasználásával, ami fontos a költségek optimalizálására és a megalapozott döntéshozatal felgyorsítására törekvő vállalatok számára, kis erőforrások felhasználásával. A következő fejezetekben a analitikára és a nyilvános adathalmazokon alapuló prediktív analitikára vonatkozó példákat vizsgálja meg a nagyadat-eszközök felhasználásával.

Nagy adatok: San Francisco egymillió építési engedélyének adatainak elemzése

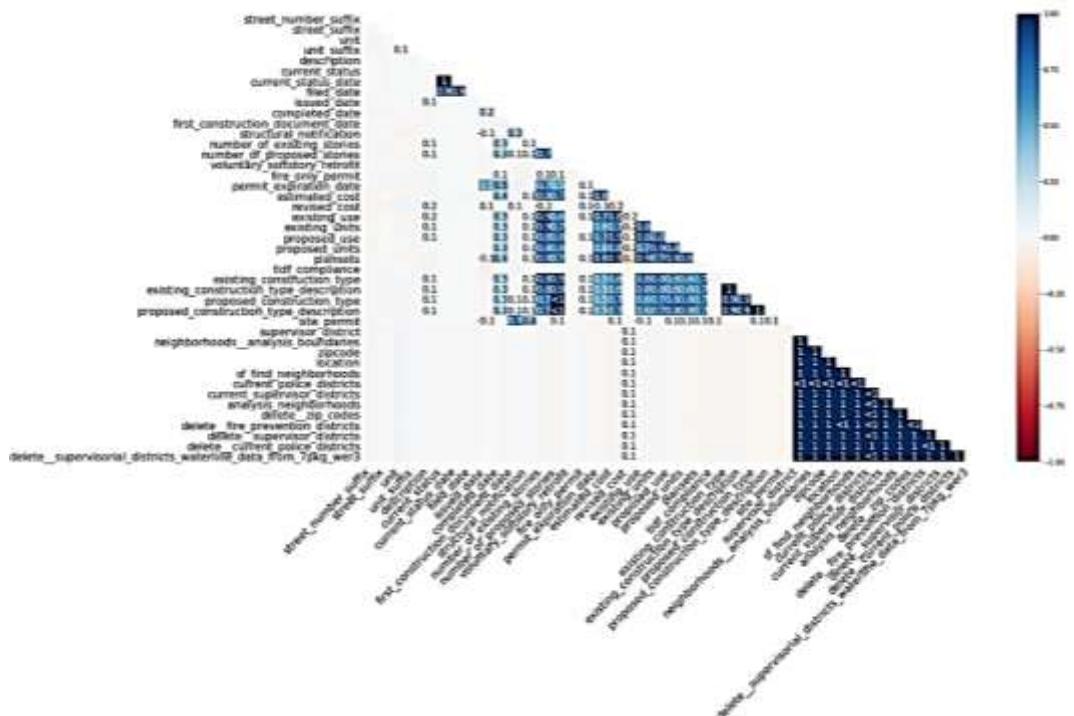
A nyílt adatkészletekkel való munka egyedülálló lehetőséget nyújt az előző fejezetekben tárgyalt elvek gyakorlati megvalósítására: a funkciók körültekintő kiválasztása, a reprezentatív mintavétel, a vizualizáció és a kritikus elemzés. Ebben a fejezetben azt vizsgáljuk meg, hogyan vizsgálhatók olyan összetett jelenségek, mint például az építési tevékenység egy nagyvárosban, nyílt adatok felhasználásával - különösen a San Franciscó-i több mint egymillió építési engedélyezési rekordot.

Nyilvánosan elérhető adatok több mint egymillió építési engedélyről (9.1-2. ábra) (két adatállományban, CSV formátumban) a "San Francisco Department of Buildings" (San Francisco Építési Főosztály) adataiból. [148] lehetővé teszik számunkra, hogy a nyers CSV -táblázatot ne csak a város építési tevékenységének elemzésére használjuk, hanem a közelmúlt tendenciáinak és San Francisco építőiparának történelmének kritikus elemzésére is az elmúlt 40 évben, 1980-tól 2019-ig.

Az adatkészlet vizualizációjának létrehozásához használt kódpéldák (9.1-3. ábra - 9.1-8. ábra), valamint a kódossal, magyarázatokkal és megjegyzésekkel ellátott viszonylagos grafikonok megtalálhatók a Kaggle platformon a "San Francisco. Construction Sector 1980-2019". [149].

count		Building Permits on or after January 1, 2013			Building Permits before January 1, 2013		
permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date	
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998	
12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004		NaN
02/18/1992	install auto fire spks.	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992	
permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location	
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94123.0	(37.7903468760499,-122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0	(37.729295851600238,-122.4644245657462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0	(37.795060025252974,-122.3953224461605)

Ábra 9.1-2 Az adatkészletek a kiadott építési engedélyekre vonatkozó információkat tartalmaznak különböző objektumattribútumokkal.

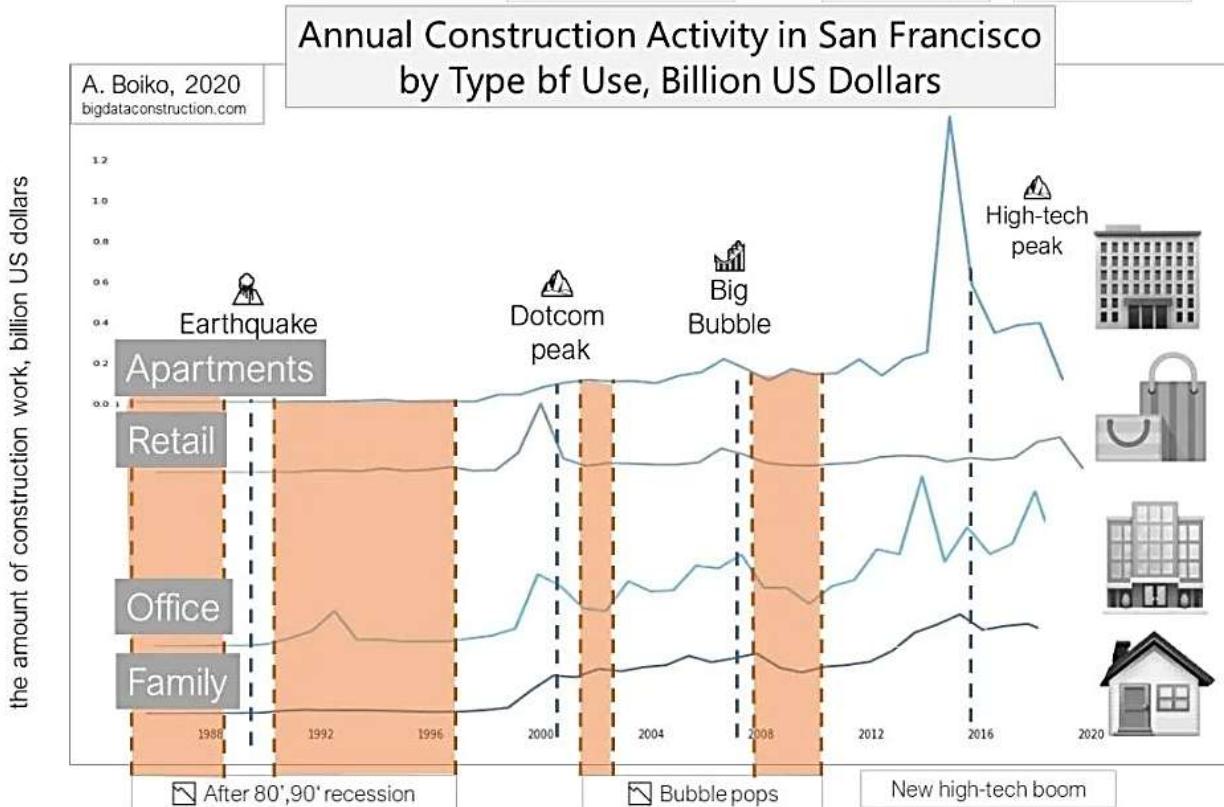
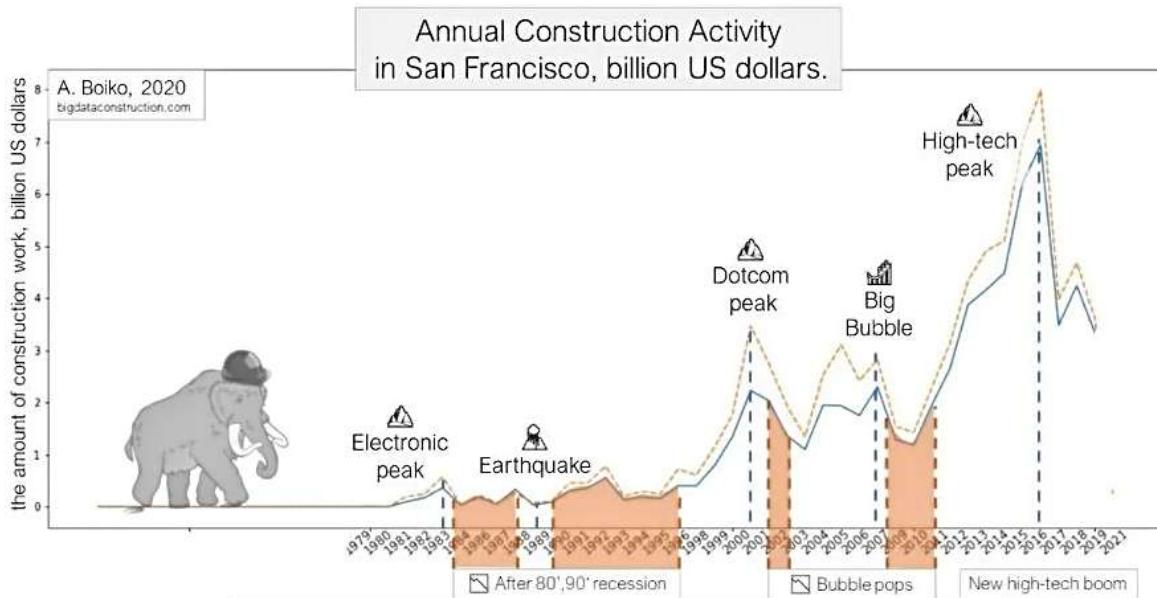


Ábra 9.1-3 Hőterkép (Pandas és Seaborn), amely egy adathalmaz összes attribútumát megjeleníti, és segít azonosítani az attribútumpárok közötti kapcsolatokat.

A San Franciscó-i Építésügyi Minisztérium által rendelkezésre bocsátott táblázatból (9.1-2. ábra) nem derül ki semmilyen tendencia vagy következetés. A táblázatos formában szereplő száraz számok nem szolgálnak döntési alapként. Ahhoz, hogy az adatok vizuálisan is érthetővé váljanak, az adatok vizualizációjáról szóló fejezetekben részletesen tárgyaltak szerint az adatokat az "ETL és az eredmények grafikonként való megjelenítése" című könyv hetedik részében tárgyalt különböző könyvtárak segítségével kell vizualizálni.

A Pandas DataFrame és a Python vizualizációs könyvtárak segítségével elemezve az 1 137 695 engedély [148] értékére adatokat, arra a következtetésre juthatunk, hogy a San Franciscó-i építési tevékenység szorosan kapcsolódik a gazdasági ciklusokhoz, különösen a Szilícium-völgy virágzó technológiai iparában (9.1-4. ábra).

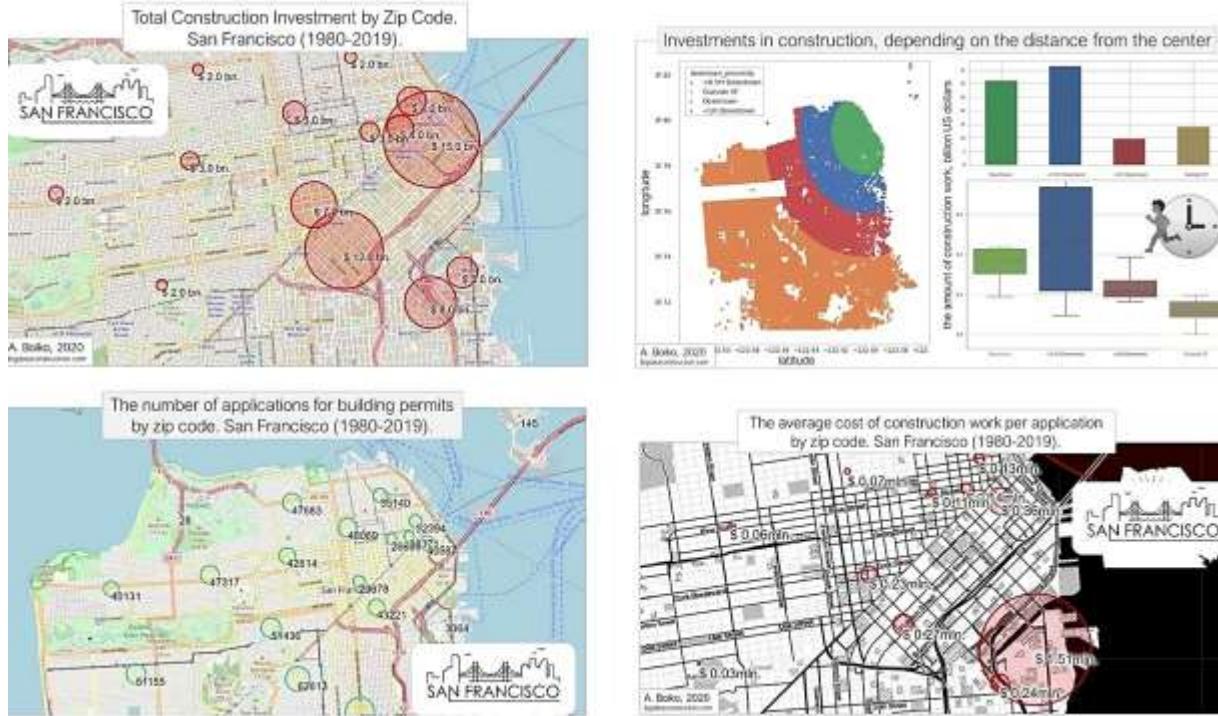
A gazdasági fellendülések és visszaesések jelentős hatással vannak az építési projektek számára és értékére. Az építési tevékenység első csúcspontja például egybeesett az 1980-as évek közepén bekövetkezett elektronikai boommal (Pandas és Matplotlib), az ezt követő csúcspontok és visszaesések pedig a dot-com buborékkal és az elmúlt évek technológiai boomjával voltak összefüggésben.



9.1-4. ábra San Francisco ingatlanszektörében a beruházások a Szilícium-völgy technológiai fejlődésével korrelálnak.

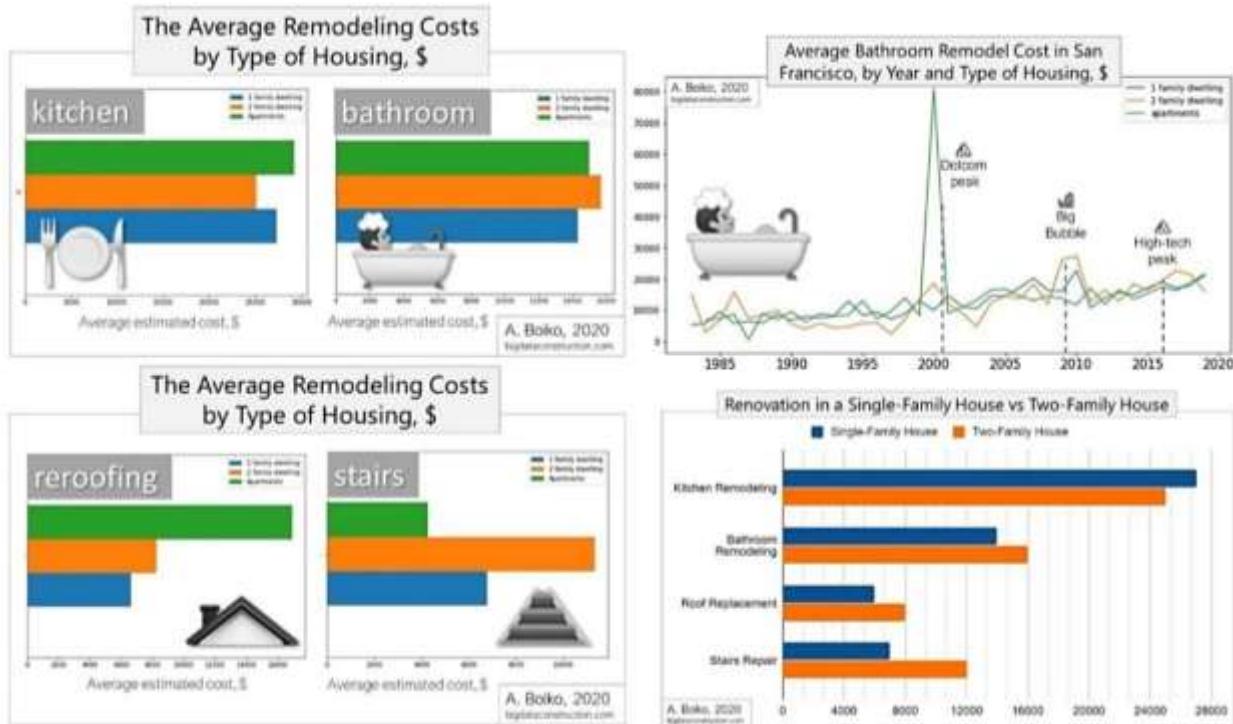
Az adatelemzés azt mutatja, hogy San Franciscóban az elmúlt évtizedben az építkezésekbe és átalakításokba befektetett 91,5 milliárd dollár nagy része - közel 75%-a - a városközpontban (9.1-5. ábra - Pandas és Folium vizualizációs könyvtár felhasználásával) és a városközpont 2 km-es körzetében összpontosul, ami a beruházások nagyobb sűrűségét tükrözi ezekben a központi övezetekben.

Az építési engedélyek átlagos költsége városrészekenként jelentősen eltér: a városközpontban a kérelmek háromszor annyiba kerülnek, mint a városközponton kívül, ami a magasabb telek-, munka- és anyagköltségeknek, valamint az energiahatékonyságot javító, drágább anyagokat megkövetelő szigorú építési előírásoknak tudható be.



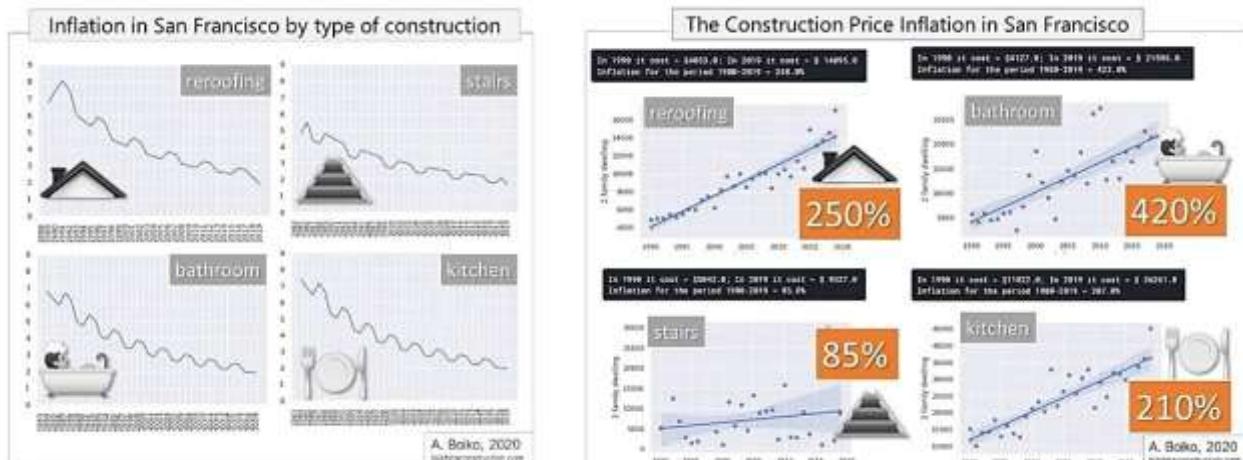
9.1-5. ábra San Franciscóban az építési beruházások 75 százaléka (91,5 milliárd dollár) a belvárosban összpontosul.

Az adatkészlet lehetővé teszi az átlagos javítási árak kiszámítását nemcsak lakástípusonként, hanem városrészekenként és egyedi címek (irányítószámok) szerint is. San Franciscóban a lakásfelújítási költségek dinamikája a különböző típusú felújítások és lakások esetében eltérő tendenciákat mutat (9.1-6. ábra - Pandas és Matplotlib). A konyha felújítása észrevehetően drágább, mint a fürdőszoba felújítása: egy átlagos konyha felújítása egy egylakásos házban körülbelül 28 000 dollárba kerül, míg egy kétlakásos házban 25 000 dollárba.



9.1-6 ábra SF-ben a konyha felújítása majdnem kétszer annyiba kerül, mint a fürdőszoba felújítása, és a lakástulajdonosoknak 15 éven keresztül havonta \$ 350 forintot kell félretenniük, hogy fedezni tudják a nagyobb lakásjavítások költségeit.

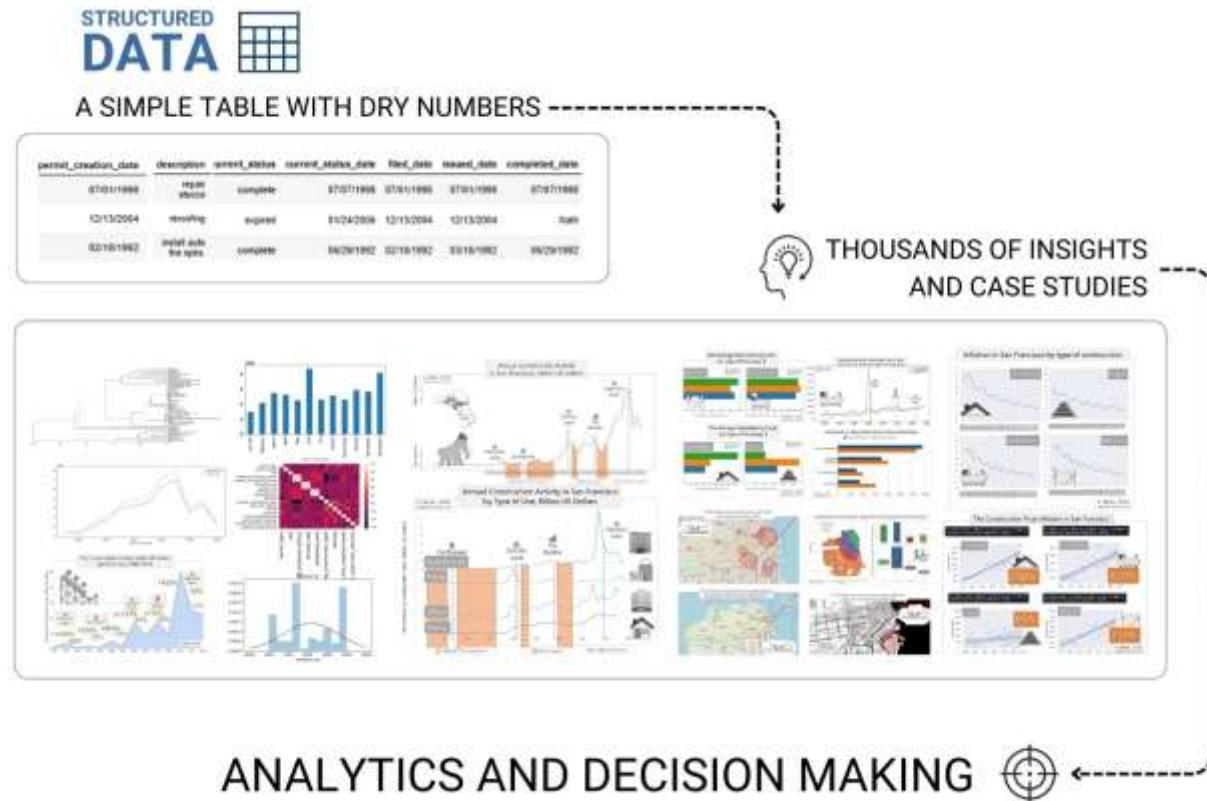
A San Franciscó-i építési költségek inflációja az évek során nyomon követhető a lakástípusok és évek szerint csoportosított adatok elemzésével (9.1-7. ábra - a Pandas és Seaborn által használt), amely az átlagos javítási költségek 1990 óta tartó folyamatos növekedését mutatja, és rövid távú hároméves ciklusokat mutat a többplakásos épületek javítási költségeiben.



9.1-7. ábra 1980 és 2019 között a fürdőszobafelújítások költségei SF-ben ötszörösére nőttek, míg a tető- és konyhafelújítások ára megháromszorozódott, a lépcsőházeljítások pedig csak 85%-kal emelkedtek.

A San Franciscói építési osztály nyilvános adatainak tanulmányozása (9.1-3. ábra) azt mutatja, hogy az építési költségek a városban rendkívül változók és gyakran kiszámíthatatlanok, és számos tényező befolyásolja őket. E tényezők közé tartozik a gazdasági növekedés, a technológiai innováció és a különböző lakástípusok egyedi követelményei.

Korábban az ilyen elemzésekhez alapos programozási és elemzési ismeretekre volt szükség. Az LLM-eszközök megjelenésével azonban a folyamat az építőiparban dolgozó szakemberek széles köre számára elérhetővé és érthetővé vált, a tervezési osztályokon dolgozó mérnököktől a felsővezetőig.



9.1-8 A vizuálisan érthető adatokra való áttérés a rejtett minták felismerésével lehetővé teszi az automatizált döntéshozatalt.

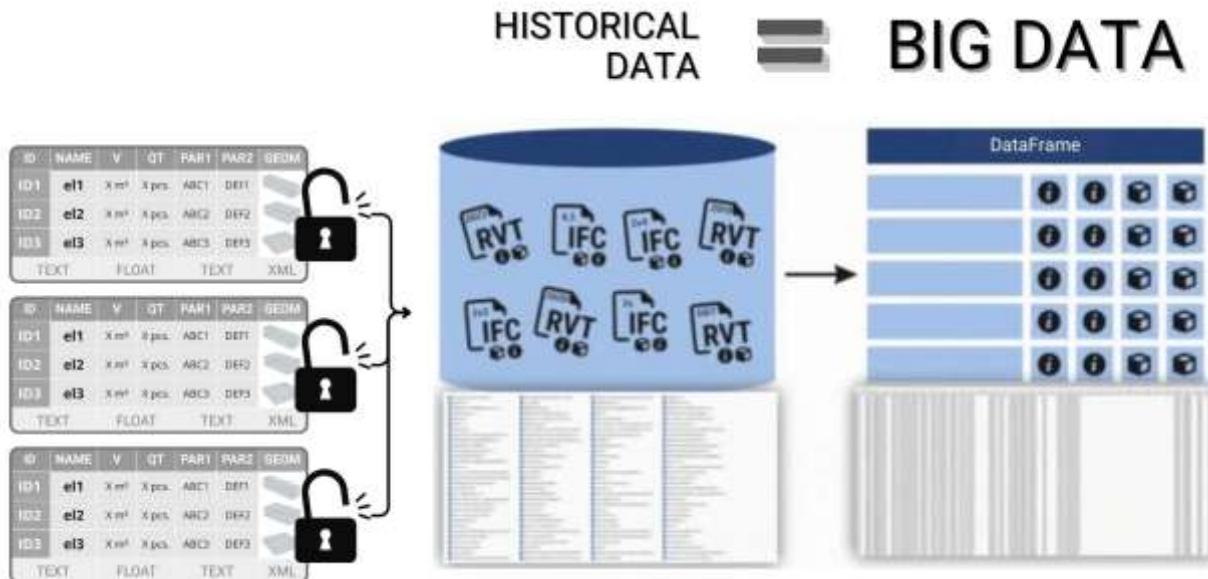
Ahogyan a "San Francisco-i építési hatóság" táblázatos adatállományának adatait elemeztük, úgy bármilyen adatállományt - a képektől és dokumentumuktól kezdve a tárgyak internetének adataiig, vagy a származtatott CAD-adatbázisok adataiig - vizualizálhatunk és elemezhetünk.

Példa a CAD-adatokon alapuló nagyméretű adatokra (BIM)

A következő példában egy nagy adathalmazt fogunk elemzni a különböző CAD-eszközökből származó adatok felhasználásával (BIM). A nagy adathalmaz összegyűjtéséhez és létrehozásához egy speciális, automatizált webkúszó programot (script) használtunk, amelyet úgy konfiguráltunk, hogy automatikusan keressen és gyűjtsön tervezési fájlokat olyan weboldalakról, amelyek ingyenes építészeti modelleket kínálnak a RVT és az IFC formátumban. A crawler néhány nap alatt sikeresen megtalált és letöltött 4596

IFC fájlt, 6471 RVT fájlt és 156 024 DWG fájlt[149].

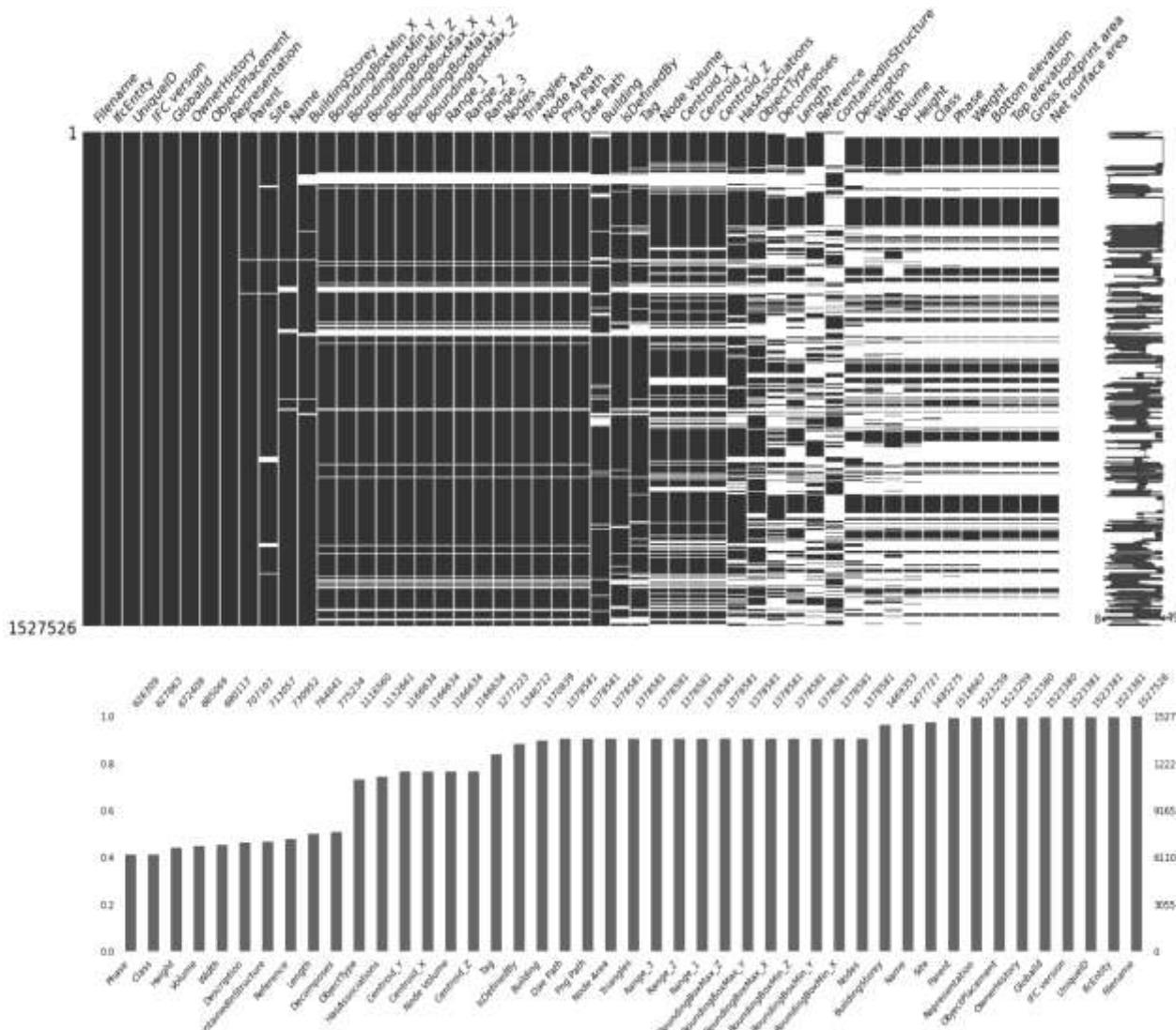
Miután összegyűjtöttük a különböző verziók RVT és IFC formátumú projektjeit, és az ingyenes reverse engineering SDK-k segítségével strukturált CSV formátumba konvertáltuk őket, közel 10 ezer RVT és IFC projektet gyűjtöttünk össze egy nagy Apache Parquet táblázatos fájlba és feltöltöttük a Pandas DataFrame programba elemzés céljából (9.1-9. ábra).



9.1-9. ábra Strukturált adatok projektadatok lehetővé teszi, hogy tetszőleges számú projektet egyetlen kétdimenziós táblázatban egyesítse.

E nagyszabású gyűjtemény adatai a következő információkat tartalmazzák: az IFC fájkészlet körülbelül 4 millió entitást (sort) és 24 962 attribútumot (oszlopot) tartalmaz, az RVT fájkészlet pedig, amely körülbelül 6 millió entitást (sort) tartalmaz, 27 025 különböző attribútumot (oszlopot) tartalmaz.

Ezek az információs halmazok (9.1-10. ábra) több millió elemet fednek le, amelyek mindegyikéhez a Bounding Box geometria (a projektben lévő objektum határait meghatározó téglalap) koordinátáit is megkaptuk, és egy közös táblázathoz adtuk - a Bounding Box geometria koordinátáit (a projektben lévő objektum határait meghatározó téglalap) és az egyes elemek PNG formátumú képeit, valamint a geometriát a - DAE (Collada) nyílt XML formátumban készítettük.



9.1-10. ábra 1,5 millió elemet tartalmazó SSubset és az első 100 attribútum foglaltságának vizualizálása (missingno könyvtár) hisztogram formájában.

Így több tízmillió elemre vonatkozó információt kaptunk 4596 IFC-projektből és 6471 RVT-projektből, ahol az összes entitáselem összes attribútuma-tulajdonsága és geometriája (Bounding Box) egyetlen táblázat (DataFrame) strukturált formájába került át (9.1-10. ábra - az adatkeret-populációk adatai hisztogramok formájában jelennek meg).

Az elemzési folyamat során kirajzolt hisztogramok (9.1-10. ábra, 9.2-6. ábra, 9.2-7. ábra) lehetővé teszik az adatsűrűség és az oszlopokban szereplő értékek előfordulási gyakoriságának gyors értékelését. Ez első pillantást ad a jellemzők eloszlására, a kiugró értékek jelenlétére és az egyes jellemzők potenciális hasznosságára a gépi tanulási modellek elemzése és felépítése során.

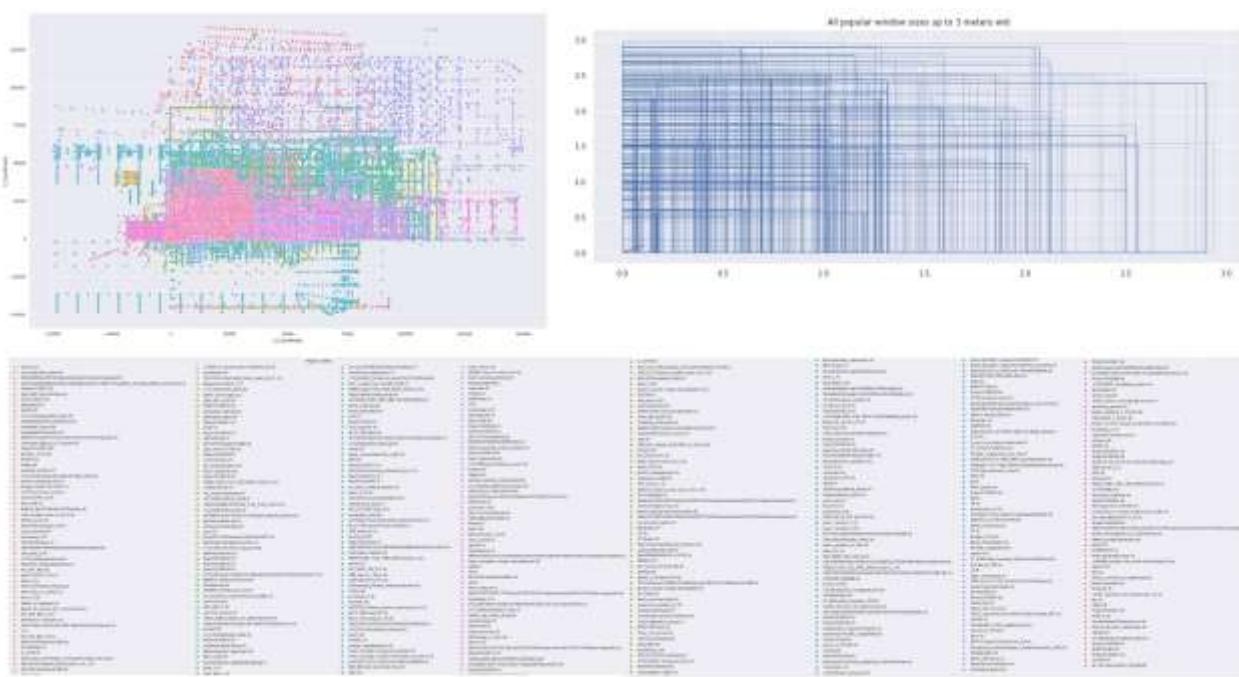
Ennek az adatkészletnek a gyakorlati felhasználására példa (9.1-10. ábra) az "5000 IFC és RVT " projekt. [149], amely a Kaggle platformon érhető el. A Jupyter Notebook teljes Pipeline-megoldást mutat be: az

adatok előfeldolgozásától és elemzésétől az eredmények vizualizálásáig Python könyvtárak - pandas, matplotlib, seaborn, folium és mások - használatával (9.1-11. ábra).



9.1-11. ábra Példák CAD-formátumokból származó adatok elemzésére (BIM) Python vizualizációs könyvtárak és a pandas könyvtár használatával.

A metainformációk alapján meghatározható, hogy mely városokban fejlesztettek ki bizonyos projekteket, és ez térképen is megjeleníthető (pl. a folium könyvtár segítségével). Ezenkívül az adatokban található időbelyegzők lehetővé teszik a fájlok mentésének vagy szerkesztésének időpontjában megfigyelhető minták feltárását: a hét napja, a napszak és a hónap szerint.



Ábra 9.1-12 Vizualizáció az összes oszlop geometriai helyzetét és az összes ablak méreteit 3 méterig a táblázat alján található listából kivétítve.

A modellekben kinyert geometriai paraméterek Bounding Box formájában szintén alkalmasak az aggregált

elemzésekre. A 9.1-12. ábra például két grafikont mutat: a bal oldali az oszlopok közötti távolságok eloszlását mutatja az összes projekt esetében a nullponthoz képest, a jobb oldali pedig az összes ablak méreteit mutatja 3 méter magasságig egy több tízezer ablakelemet tartalmazó mintában (miután a teljes adathalmazt a "Category" paraméter "OST_Windows", "IfcWindows" értékkel csoportosítottuk).

A példa Pipeline-elemzési kódja és maga az adathalmaz elérhető a Kaggle weboldalán "5000 IFC és RVT | DataDrivenConstruction.io projects" [149] címen. Ez a kész Pipeline az adatkészlettel együtt lemagasoltató és futtatható ingyenesen online a Kaggle oldalon vagy offline a népszerű IDE-k valamelyikében: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse a PyDev pluginnal, Thonny, Wing IDE, IntelliJ IDEA a Python pluginnal, JupyterLab vagy a népszerű online eszközök Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

A hatalmas mennyiségű strukturált adat feldolgozásából és tanulmányozásából nyert elemzői meglátások döntő szerepet fognak játszani az építőipar döntéshozatali folyamataiban.

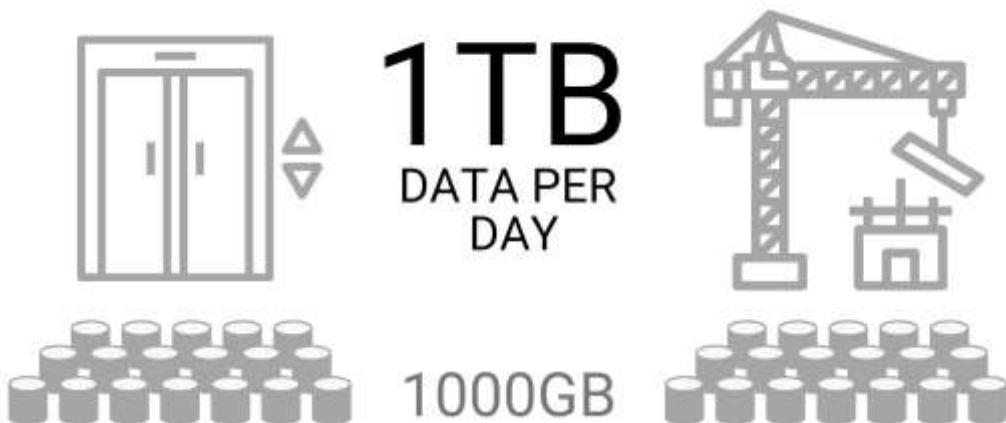
A szakemberek a múltbeli projektek alapján végzett ilyen jellegű információelemzéssel hatékonyan előre tudják jelezni például az anyag- és munkaerő-szükségletet, és még az építkezés megkezdése előtt optimalizálni tudják a tervezési megoldásokat.

Míg azonban a tervezési adatok vagy az építési engedélyek viszonylag statikus információk, amelyek viszonylag lassan változnak, maga az építési folyamat gyorsan telítődik különféle érzékelőkkel és IoT - eszközökkel: kamerák, automatizált megfigyelőrendszerek, amelyek valós időben továbbítják az adatokat - mindenek dinamikus digitális környezetté teszik az építkezési területet, ahol az adatokat valós időben kell elemezni.

IoT A dolgok internete és az intelligens szerződések

IoT A tárgyak internete a digitális átalakulás új hullámát jelenti, amelyben minden eszköz saját IP-címet kap, és egy globális hálózat részévé válik. Az IoT egy olyan koncepció, amely fizikai tárgyak internethoz való csatlakoztatását foglalja magában az adatok gyűjtése, feldolgozása és továbbítása érdekében. Az építőiparban ez az építési folyamatok valós idejű vezérlését, az anyagpazarlás minimalizálását, a berendezések elhasználódásának előrejelzését és a döntéshozatal automatizálását jelenti.

A CFMA "Felkészülés a jövőre az összekapcsolt építéssel" című cikke szerint [150] szerint az építőipar a következő évtizedben jelentős digitális átalakuláson megy keresztül, amely a Connected Construction koncepcióban - egy teljesen integrált és automatizált építkezésben - csúcsosodik ki.



Ábra 9.1-13 Az IOT vagy az építkezés adatátviteli eszközei naponta terabajtnyi adatot állíthatnak elő és továbbíthatnak.

A digitális építkezés azt jelenti, hogy az építkezés minden eleme - a tervezéstől és a logisztikától a munka kivitelezéséig és a minőségellenőrzésig - egyetlen dinamikus digitális ökoszisztemába kerül integrálásra. Korábban, e könyv 7. részében már áttekintettük az Apache NiFi (7.4-5. ábra) képességeit, egy olyan ingyenes és nyílt forráskódú eszközöt, amely lehetővé teszi a valós idejű adatáramlást - a különböző forrásokból történő gyűjtéstől a tárolási vagy elemzési platformokra történő átvitelig.

Az építkezés előrehaladásáról, az anyagfelhasználásról, a berendezések állapotáról és a biztonságról szóló adatokat valós időben továbbítják az elemző rendszerekhez (9.1-13. ábra). Ez lehetővé teszi a lehetséges kockázatok előrejelzését, az eltérésekre való azonnali reagálást és a helyszíni folyamatok optimalizálását. A digitális építkezés kulcsfontosságú elemei a következők:

- IoT - érzékelők - a környezeti paraméterek nyomon követése, az építőipari berendezések felügyelete és a munkafeltételek ellenőrzése.
- Digitális iker - az épületek és az infrastruktúra virtuális modelljei az esetleges eltérések előrejelzésére és a hibák megelőzésére.
- Automatizált logisztikai rendszerek - valós idejű ellátási lánc menedzsment az állásidő és a költségek csökkentése érdekében.
- Robotizált építőipari rendszerek - autonóm gépek alkalmazása rutinszerű és veszélyes feladatok elvégzésére.

A robotizáció, az IoTHasználata és a Connected Site (építőipar) digitális építkezés koncepciója nemcsak a hatékonyságot növeli és a költségeket csökkenti, hanem a biztonság, a fenntartható építés és a prediktív projektmenedzsment új korszakát is megnyitja.

Az RFID (rádiófrekvenciás azonosító) címkék szintén az IoT egyik legfontosabb elemei. Ezeket az építkezéseken az anyagok, gépek és akár a személyzet azonosítására és nyomon követésére használják, növelte a projekt erőforrásainak átláthatóságát és ellenőrzését.

Az RFID -technológiát a tárgyak automatikus felismerésére használják rádiójelek segítségével. Hárrom kulcsfontosságú elemből áll:

- RFID - (passzív vagy aktív) címkék - egyedi azonosítót tartalmaznak, és az anyagokhoz,

eszközökhöz vagy gépekhez rögzítik őket.

- A szkennerek olyan eszközök, amelyek információkat olvasnak le a címkékről, és továbbítják azokat a rendszerbe.
- Központi adatbázis - tárolja a tárgyak helyére, állapotára és mozgására vonatkozó információkat.

Az RFID alkalmazása az építőiparban:

- Automatikus anyagnyilvántartás - a transzportbeton termékeken, a betoncélokban vagy a szendvicspanel-csomagokon elhelyezett címkék lehetővé teszik a készletellenőrzést és megakadályozzák a lopásokat.
- Személyzeti munka ellenőrzése - RFID - a munkavállalói jelvények rögzítik a műszakkezdés és - zárás időpontját, így a munkaórák nyilvántartását.
- Berendezésfigyelés - RFID - rendszer követi a berendezések mozgását, megelőzve az állásidőt és javítva a logisztikai hatékonyságot.

Ezt a technológiai csomagot blokklánc-alapú intelligens szerződések egészítik ki, amelyek közvetítők nélkül automatizálják a kifizetéseket, a szállítás ellenőrzését és a szerződéseknek való megfelelést, csökkentve a csalás és a késedelmek kockázatát.

Ma, közös adatmodell hiányában az intelligens szerződések egyszerűen csak kódok, amelyekben a résztvevők megállapodnak. Az adatközpontú megközelítéssel azonban lehetséges a szerződés paramétereinek közös modelljét létrehozni, azt egy blokkláncba kódolni, és automatizálni a feltételek teljesítését.

Például egy ellátási lánc menedzsment rendszerben egy intelligens szerződés képes nyomon követni egy szállítmány kézbesítését az IoT -érzékelők és az RFID -címkék segítségével, és automatikusan átutalni a fizetést, amikor az megérkezik. Hasonlóképpen, egy építkezésen egy intelligens szerződés rögzítheti egy munkafázis - például a betonvasak felszerelése vagy az alapozás - befejezését a drónok vagy építési érzékelők adatai alapján, és automatikusan kezdeményezheti a következő kifizetést a vállalkozónak, kézi ellenőrzések és papíralapú igazolások nélkül.

Az új technológiák és a nemzetközi szabványügyi szervezetek erőfeszítései ellenére azonban az egymással versengő szabványok sokasága bonyolítja a tárgyak internetét.

A Cisco 2017-ben közzétett tanulmánya [151] szerint a tárgyak internetére vonatkozó kezdeményezések (IoT) közel 60%-a megáll a koncepció bizonyítási szakaszában, és a vállalatoknak csak 26%-a tartja teljesen sikeresnek IoT-projektjét. Ráadásul a befejezett projektek egyharmada nem éri el a kitűzött célokat, és a megvalósítás után sem ismerik el őket sikeresnek.

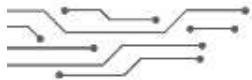
Ennek egyik fő oka a különböző érzékelők adatait feldolgozó platformok közötti átválasztás hiánya. Ennek eredményeképpen az adatok különálló megoldásokon belül maradnak elszigetelve. Ennek a megközelítésnek az alternatívája, ahogy más hasonló esetekben is (amelyekkel ebben a könyvben foglalkoztunk), egy olyan architektúra, amely magára az adatra, mint elsődleges eszközre épül.

Az IoT-érzékelők nemcsak a berendezések műszaki állapotának nyomon követésében játszanak kulcsszerepet, hanem a prediktív elemzésben is, hogy a meghibásodások és eltérések előrejelzésével csökkentsék a kockázatokat az építkezésen, és javítsák a folyamatok általános teljesítményét.

Az IoT-érzékelők és RFID-címkek által gyűjtött adatok valós időben feldolgozhatók gépi tanulási algoritmusokkal, amelyek képesek észlelni a rendellenességeket, és előre figyelmeztetni a mérnököket a lehetséges meghibásodásokra. Ez a betonszerkezetek mikrorepedéseitől kezdve a toronydaru működésének jellegtelennel szüneteiig terjedhet, amelyek műszaki hibákra vagy a jogszabályok megsértésére utalnak. A fejlett viselkedéselemző algoritmusok továbbá képesek olyan viselkedési mintákat rögzíteni, amelyek például a személyzet fizikai fáradtságára utalhatnak, így fokozva a biztonság és a munkavállalók jólétének proaktív kezelését a helyszínen.

Az építőiparban a balesetek és meghibásodások - legyen szó akár gépekről, akár emberekről - ritkán történnek hirtelen. Általában kisebb eltérések előzik meg őket, amelyek észrevétlenül maradnak. A prediktív analitika és a gépi tanulás lehetővé teszi, hogy ezeket a jeleket korai szakaszban, még a kritikus következmények bekövetkezése előtt észleljük.

Míg a dokumentumok, a projektfájlok és az IoT-eszközök és RFID-címkek származó adatok alkotják az építési projektek digitális lábnyomát, a gépi tanulás segíthet hasznos információkat kinyerni belőlük. Az adatok növekedésével és az adatokhoz való hozzáférés demokratizálódásával az építőipar új lehetőségeket nyer az analitika, a prediktív analitika és a mesterséges intelligencia alkalmazásai terén.

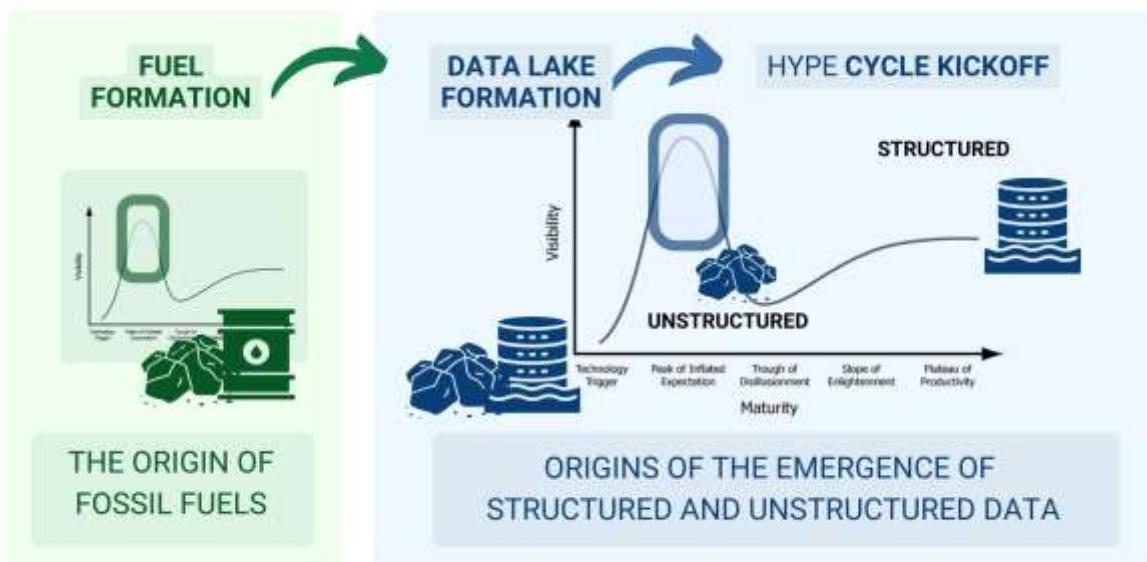


FEJEZET 9.2.

GÉPI TANULÁS ÉS ELŐREJELZÉSEK

A gépi tanulás és a mesterséges intelligencia meg fogja változtatni az építkezésünket.

Az építőipar különböző rendszereinek adatbázisai - elkerülhetetlenül hanyatló és egyre bonyolultabb infrastruktúrájukkal - a jövőbeli megoldások táptalajává válnak. A vállalati szerverek, akár egy erdő, fontos információk biomasszájával gazdagodnak, amelyek gyakran a föld alatt, a mappák és szerverek gyomrában rejtőznek. A különböző rendszerekből ma keletkező adattömegek - a használat után, a szerver aljára hullva és évekig tartó megkövesedés után - a jövőben a gépi tanulást és a nyelvi modellek fogják táplálni. A belső vállalati csevegőszobák (pl. a helyileg konfigurált ChatGPT, LlaMa, Mistral, DeepSeek külön példánya) ezekre a házon belüli modellekre épülnek majd, a központosított tárolás segítségével, hogy gyorsan és kényelmesen lekérdezzék az információkat, és létrehozzák a szükséges grafikonokat, műszerfalakat és dokumentumokat.



Ábra 9.2-1 Ahogy a fákból szén lesz, úgy az információ is értékes üzleti energiává alakul át idővel az idő és az analitika nyomása alatt.

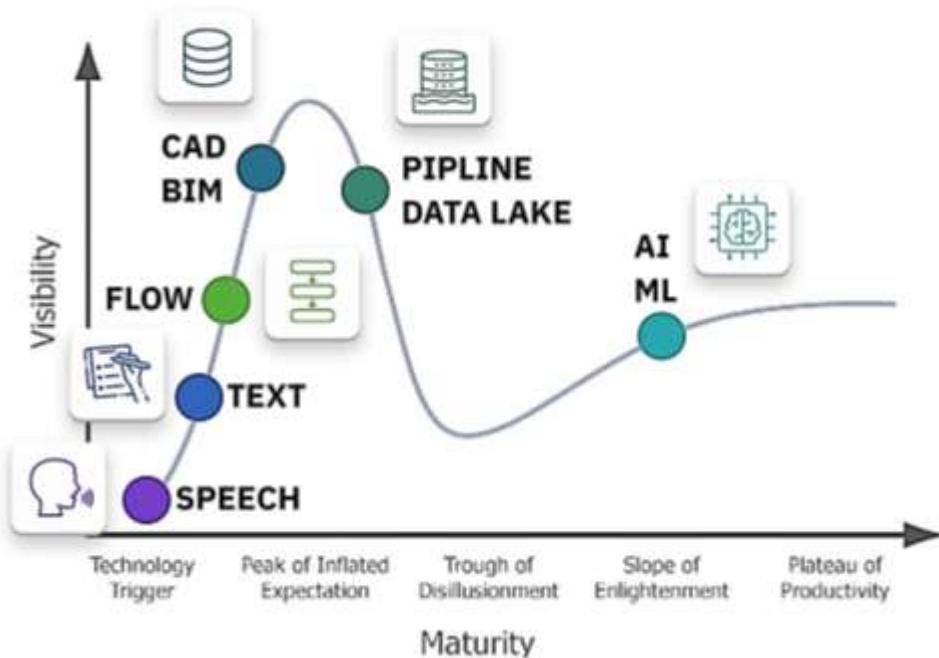
A növényi tömeg megkövesedése a nyomással és a hőmérséklettel együttesen homogén és egyedi szerkezetű, különböző időpontokban élt, különböző fajokból származó fák homogén tömegét - a faszen [152] - hozza létre. Ugyanígy a merevlemezekre különböző formátumokban és különböző időpontokban rögzített információk az analitikai osztályok nyomása és a minőségirányítás hőmérséklete alatt végül az értékes információk homogén szerkezetű, strukturált tömegét alkotják (9.2-1. ábra).

Ezeket az információs rétegeket (vagy gyakrabban elszigetelt rögöket) a tapasztalt elemzők aprólékos adatszervezésével hozzák létre, akik fokozatosan kezdkik kivonni a értékes információkat a látszólag hosszú ideig irreleváns adatokból.

Abban a pillanatban, amikor ezek az érett adatrétegek már nem csak jelentésekben "égnek", hanem

elkezdenek az üzleti folyamatokban keringeni, gazdagítják a döntéseket és javítják a folyamatokat, a vállalat készen áll a következő lépésre - a gépi tanulásra és a mesterséges intelligenciára való áttérésre (9.2-2. ábra).

A gépi tanulás (ML - Machine learning) a mesterséges intelligencia problémák megoldására szolgáló módszerek egy osztálya. A gépi tanulási algoritmusok nagy adathalmazokban felismerik a mintákat, és ezek alapján tanulnak. minden egyes új adathalmaz lehetővé teszi a matematikai algoritmusok számára, hogy a kapott információknak megfelelően javuljanak és alkalmazkodjanak, ami lehetővé teszi az ajánlások és előrejelzések pontosságának folyamatos javítását.



Ábra 9.2-2 Az adatalkotási technológiák elhalványulása és az analitikai eszközök alkalmazása megnyitja az ajtót a gépi tanulás témája előtt.

Ahogy a világ legnagyobb befektetési alapjának befolyásos vezérigazgatója (amely kulcsfontosságú részesedésekkel rendelkezik szinte az összes legnagyobb építőipari szoftvercégben, valamint a világ legnagyobb ingatlanvagyonát birtokló vállalatokban [55]) egy 2023-as interjúban elmondta. - a gépi tanulás meg fogja változtatni az építőipar világát.

A mesterséges intelligencia óriási lehetőségeket rejt magában. Meg fogja változtatni a munkánkat és az életünket. A mesterséges intelligencia és a robotika meg fogja változtatni a munkánkat és az építkezésünket, és képesek leszünk a mesterséges intelligenciát és a robotikát eszközként használni a sokkal nagyobb termelékenység megteremtésére [153].

- A világ legnagyobb befektetési alapjának vezérigazgatója, interjú, 2023. szeptember.

A gépi tanulás (ML) nagy mennyiségű adat feldolgozásával működik, statisztikai technikákat alkalmazva az emberi gondolkodás egyes aspektusainak utánzására. A legtöbb vállalat azonban nem rendelkezik ilyen adatkészletekkel, és ha mégis, akkor azok gyakran nincsenek kellően címkézve. Ebben segíthetnek a szemantikus technológiák és a transzfer tanulás, egy olyan technika, amely lehetővé teszi, hogy az ML hatékonyabb legyen kis mennyiségű adat feldolgozásakor, és amelynek megvalósíthatóságát e rész korábbi fejezeteiben már tárgyaltuk.

A transzfer tanulás lényege, hogy ahelyett, hogy minden feladatot a nulláról dolgozna fel, felhasználhatja a kapcsolódó területeken szerzett tudást. Fel kell ismerni, hogy a más iparágakból származó minták és felfedezések adaptálhatók és alkalmazhatók az építőiparban. Például a kiskereskedelemben kifejlesztett logisztikai folyamatok optimalizálásának módszerei segítenek az építőipari ellátási lánc menedzsment hatékonyságának javításában. A pénzügyekben aktívan alkalmazott nagy adatelemzés alkalmazható az építési projektek költségelőrejelzésére és kockázatkezelésére. Az iparban kifejlesztett számítógépes látás- és robotikai technológiák pedig már alkalmazásra találnak az automatizált minőségellenőrzésben, a biztonsági felügyeletben és az építkezési létesítmények irányításában.

A transzfer-tanulás nemcsak az innovációk bevezetésének felgyorsítását teszi lehetővé, hanem azt is, hogy más iparágak már felhalmozott tapasztalatait felhasználva csökkentse a fejlesztés költségeit.

$$\text{labor productivity in construction} = f(\text{AI})$$

Ábra 9.2-3 Mesterséges intelligencia technológiák és a robotika lesz a jövő fő hajtóereje az építőipar termelékenységének növelésében.

Az emberi gondolkodás hasonló elven szerveződik: a korábban megszerzett tudásra építünk, hogy új problémákat oldjunk meg (4.4-19. ábra, 4.4-20. ábra, 4.4-21. ábra). A gépi tanulásban is működik ez a megközelítés - az adatmodell egyszerűsítésével és elegánsabbá tételevel csökkenthetjük a probléma komplexitását az ML-algoritmusok számára. Ez pedig csökkenti a nagy mennyiségű adat szükségességét és a számítási költségeket.

A szubjektív értékeléstől a statisztikai előrejelzésig

Az a korszak, amikor a stratégiai döntések az egyes vezetők intuíciójától függtek (9.2-4. ábra), már a múlté. Az egyre növekvő versenyben és kihívásokkal teli gazdasági környezetben a szubjektív megközelítés túlságosan kockázatossá és hatástartanná válik. Azok a vállalatok, amelyek a adatok objektív elemzése helyett továbbra is a személyes véleményekre hagyatkoznak, elveszítik a változásra való gyors reagálás képességét.

A versenykörnyezet megköveteli az adatokon, statisztikai mintákon és kiszámítható valószínűségen alapuló pontosságot és megismételhetőséget. A döntések többé nem alapulhatnak érzésen, hanem az analitikából és a gépi tanulásból származó összefüggéseken, trendeken és előrejelző modelleken kell alapulniuk. Ez nem csupán az eszközökben bekövetkező változás - ez a gondolkodás logikájának megváltozása: a feltételezésekkel a bizonyítékokra, a szubjektív valószínűségekről a statisztikailag kiszámított eltérésekre, az érzésekkel a tényekre.



Ábra 9.2-4 A HiPPO (a legjobban fizetett alkalmazott véleménye) által hozott döntések kora a nagy adatok és a gépi tanulás megjelenésével a múlté lesz.

Azok a vezetők, akik korábban kizártak a saját érzéseikre hagyatkoztak, elkerülhetetlenül szembesülnek az új valósággal: a tekintély már nem határozza meg a döntéseket. A menedzsment középpontjában ma már olyan rendszerek állnak, amelyek paraméterek és vektorok millióit elemzik, rejtett mintákat azonosítanak és optimális stratégiákat javasolnak.

A fő ok, amiért a vállalatok ma még mindig kerülik az ML bevezetését, az átláthatóság hiánya. A legtöbb modell "fekete dobozként" működik a vezetők számára, anélkül, hogy megmagyaráznák, hogy pontosan hogyan jutnak el a következtetésekhez. Ez problémákhöz vezet: az algoritmusok megerősíthetik a sztereotípiákat, és akár humoros helyzeteket is teremthetnek, mint például a Microsoft chatbotja esetében, amely gyorsan mérgező kommunikációs eszközzé vált [154].

A Deep Thinking című könyvben Garri Kaszparov, a korábbi sakkkvilágbajnok az IBM Big Blue számítógépéptől elszenvedett vereségéről elmélkedik [155]. Azzal érvel, hogy az AI igazi értéke nem az emberi intelligencia másolásában, hanem a képességeink kiegészítésében rejlik. Az AI-nak olyan feladatokat kell elvégeznie, amelyekben az ember gyenge, míg az ember kreativitást hoz. A számítógépek megváltoztatták a sakk elemzésének hagyományos megközelítését. Ahelyett, hogy lenyűgöző történeteket írnának a játszmákról, a számítógépes sakkprogramok minden egyes lépést páratlanul, kizárolag a tényleges erősség vagy gyengeség alapján értékelnek. Kaszparov megjegyzi, hogy az az emberi hajlam, hogy az eseményeket összefüggő történeteknek, nem pedig egyéni lépéseknek tekintsük, gyakran vezet téves következtetésekhez - nemcsak a sakkban, hanem általában az életben is.

Ezért ha azt tervezи, hogy a gépi tanulást előrejelzésre és elemzésre használja, fontos, hogy megértse az alapelveket - az algoritmusok működését és az adatok feldolgozását -, mielőtt elkezdi használni a gépi tanulási eszközöket és az AI a munkájában. A legjobb módja a kezdésnek a gyakorlati tapasztalatszerzés.

A gépi tanulás és előrejelzés téma körébe való kezdeti bevezetéshez az egyik legmegfelelőbb eszköz a Jupyter Notebook és a népszerű klasszikus Titanic adatkészlet, amely vizuális bevezetést nyújt az adatok elemzésének és az ML modellek építésének legfontosabb módszereire.

Titáni adatkészlet: Hello World az analitika világában adatok és nagy adatok

Az ML adatelemzésben való felhasználásának egyik leghíresebb példája a Titanic adathalmazának elemzése, amelyet gyakran használnak az utasok túlélési valószínűségének vizsgálatára. Ennek a táblázatnak a megtanulása analóg a programozási nyelvek tanulásakor a "Hello World" programmal.

Az RMS Titanic 1912-es elsüllyedése 1502 ember halálát okozta a 2224 emberből. A Titanic-adatkészlet nemcsak arra vonatkozó információkat tartalmaz, hogy egy utas túlélte-e, hanem olyan attribútumokat is, mint például: életkor, nem, jegyosztály és egyéb paraméterek. Ez az adatkészlet ingyenesen elérhető, és különböző offline és online platformokon megnyitható és elemezhető.

Link a Titanic-adatkészlethez:

<https://raw.githubusercontent.com/datasets/master/titanic.csv>

Korábban az "LLM-kompatibilis IDE-k és a jövőbeli programozási változások" című fejezetben már tárgyalunk a Jupyter Notebook - az adatelemzés és a gépi tanulás egyik legnépszerűbb fejlesztői környezetét. A Jupyter Notebook ingyenes felhőalapú analógai a Kaggle és a Google Collab platformok, amelyek lehetővé teszik a Python kód futtatását szoftver telepítése nélkül, és ingyenes hozzáférést biztosítanak a számítási erőforrásokhoz.

A Kaggle a legnagyobb adatelemzési, gépi tanulási versenyplatform, integrált kódfuttatási környezettel. A Kaggle-nek 2023 októberében több mint 15 millió felhasználója volt [156] 194 országból.

Tölts le és használja a Titanic adathalmazt a Kaggle platformon (9.2-5. ábra), hogy tárolja az adathalmazt (annak egy másolatát), és futtassa a Python kódot az előre telepített könyvtárakkal közvetlenül egy

böngészőben, anélkül, hogy külön IDE-t kellene telepítenie.



Ábra 9.2-5 Titanic táblázat statisztikái - az adatelemzés és a gépi tanulás tanulásának legnépszerűbb gyakorló adathalmaza.

A Titanic-adatkészlet az RMS *Titanic* fedélzetén az 1912-es hajótörés idején tartózkodó 2224 utas adatait tartalmazza. Az adatkészlet két különálló táblázatban, egy gyakorló (train.csv) és egy teszt (test.csv) mintaként van bemutatva, így egyaránt használható a modellek kiképzésére és pontosságuk értékelésére új adatokon.

A képzési adathalmaz tartalmazza az utasok attribútumait (életkor, nem, jegyosztály és egyéb), valamint a túlélőkre vonatkozó információkat (a "Túlélő" bináris értékű oszlop). A képzési adatállományt (9.2-6. ábra - train.csv fájl) a modell képzéséhez használjuk. A tesztadatkészlet (9.2-7. ábra - test.csv fájl) csak az utasok attribútumait tartalmazza a túlélőre vonatkozó információk nélkül (egyetlen "Survivor" oszlop nélkül). A tesztadatkészlet a modell új adatokon történő tesztelésére és pontosságának értékelésére szolgál.

Így a gyakorló és a tesztadathalmazban az utasok attribútumai majdnem azonosak. Az egyetlen lényeges különbség az, hogy a tesztadatkészletben olyan utasok listája található, akik nem rendelkeznek a "Survivor" oszloppal - a célváltozóval, amelyet különböző matematikai algoritmusok segítségével szeretnénk megtanulni megjósolni. A modell felépítése után pedig össze tudjuk majd hasonlítani a modellünk kimenetét a tesztadathalmazból származó valós "Survivor" paraméterrel, amelyet figyelembe veszünk az eredmények értékeléséhez.

A táblázat fő oszlopai, az utasok paraméterei a képzési és a tesztadathalmazban:

- **PassengerId** - az utas egyedi azonosítója
- **Túlélő** - 1, ha az utas túlélte, 0, ha meghalt (a tesztkészletben nem áll rendelkezésre).
- **Pclass** - jegyosztály (1, 2 vagy 3)
- **Név** - az utas neve

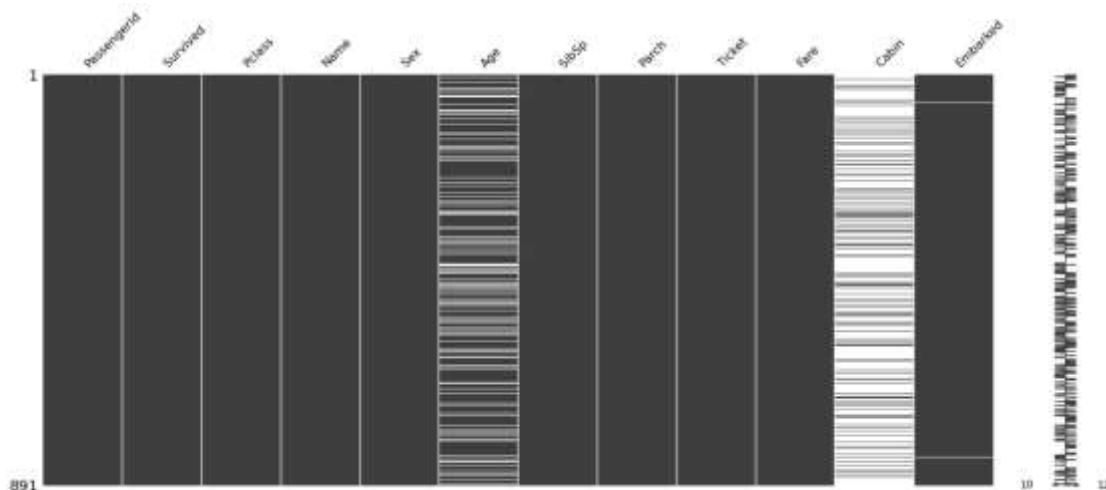
- **Nem** - az utas neme (férfi/nő)
- **Kor**
- **SibSp** - a fedélzeten tartózkodó testvérek vagy házastársak száma
- **Parch** - a szülők vagy gyermekek száma a fedélzeten
- **Jegy** - jegy száma
- **Viteldíj** - jegyár
- **Kabin** - kabinszám (sok adat hiányzik)
- **Az Embarked** a beszállás kikötője (C = Cherbourg, Q = Queenstown, S = Southampton).

A hiányzó adatok megjelenítéséhez minden két táblázatban használhatja a missingno könyvtárat (9.2-6. ábra, 9.2-7. ábra), amely a hiányzó értékeket hisztogram formájában jeleníti meg, ahol a fehér mezők jelzik a hiányzó adatokat. Ez a megjelenítés lehetővé teszi az adatok minőségének gyors értékelését a feldolgozás előtt.

```

1 import pandas as pd
2 import missingno as msno
3 import matplotlib.pyplot as plt
4
5 # Load the dataset
6 train = pd.read_csv('train.csv')
7
8 # Visualize missing values
9 msno.matrix(train)
10 plt.show()

```



Ábra 9.2-6 Néhány sornyai kódot használunk a hiányzó adatok megjelenítésére a Titanic képzési adathalmazban, ahol a képzés kulcsparamétere a "Survived" paraméter.



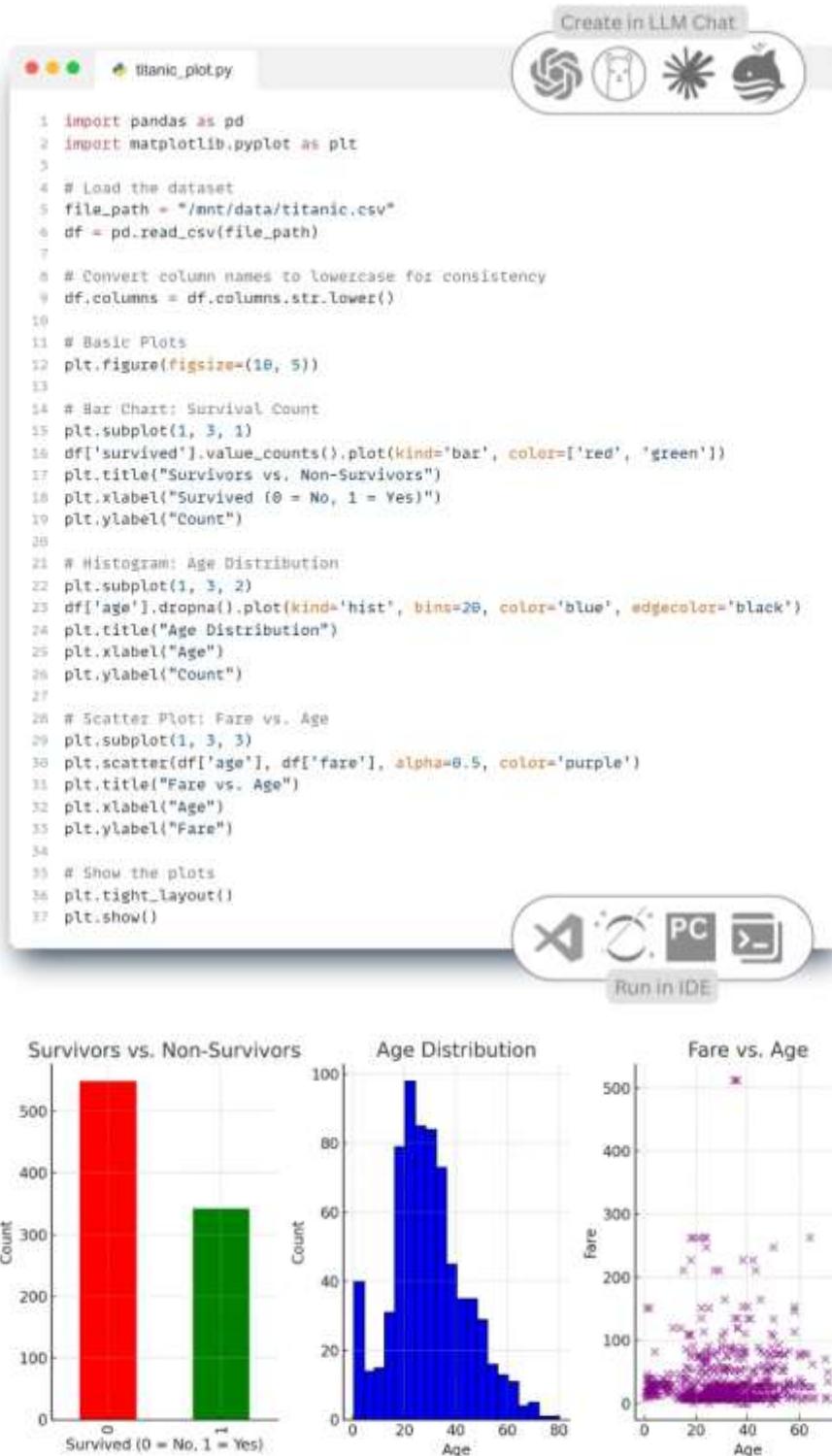
9.2-7. ábra Vizualizáció hiányzó adatok a Titanic tesztadathalmazban, amely csak az utasok jellemzőit tartalmazza információ nélkül.

Az adathalmazon alapuló hipotézisek megfogalmazása és előrejelzések készítése előtt a vizuális elemzések segítenek azonosítani a kulcsfontosságú mintákat az adatokban, értékelni azok minőségét és azonosítani a lehetséges függősségeket. Számos olyan vizualizációs technika létezik, amely segíthet a Titanic-adatkészlet jobb megértésében. Használhat eloszlási ábrákat az utasok korcsoportjainak elemzéséhez, túlélési diagramokat nemek és osztályok szerint, valamint hiányzó adatmátrixokat az információk minőségének értékeléséhez és az adatok megértéséhez.

- ❷ Kérjük meg az LLM programot, hogy segítsen nekünk a Titanic adathalmaz adatainak vizualizálásában az alábbi szöveges kérés elküldésével bármelyik LLM modellnek (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN vagy bármely más):

Kérem, mutasson néhány egyszerű grafikont a Titanic-adatkészlethez. Töltsé le maga az adathalmazt, és mutassa meg a. ↴

- LLM-válasz kész kód és az adatkészlet paramétereit megjelenítő grafikonok formájában



9.2-8. ábra LLM segítségével azonnal megjelenítheti az adatkészlet adatait.

Az adatok vizualizálása fontos lépés az adathalmaz előkészítésében a gépi tanulási modell későbbi felépítéséhez, amely csak az adatok megértésével érhető el.

Gépi tanulás akcióban: a Titanic utasaitól a projektmenedzsmentig

A fő hipotézis, amelyet a gépi tanulási keretrendszer feltárásához használtak a Titanic adathalmaz alapján, az, hogy az utasok bizonyos csoportjainak nagyobb esélyük volt a túlélésre.

A Titanic utasainak kis táblázata világszerte népszerűvé vált, és emberek milliói használják azt a Titanic utasainak képzési adathalmaza alapján képzésre, kísérletezésre és modelltesztelésre, hogy kiderítsék, mely algoritmusok és hipotézisek alkotják a leg pontosabb túlélési előrejelző modellt.

A Titanic-adatkészlet vonzereje a tömörségéből adódik: több száz sor és tizenkét oszlop (9.2-6. ábra) bőséges lehetőséget biztosít az elemzésre. Az adatkészlet viszonylag egyszerűen a bináris osztályozási megoldás klasszikus példája, ahol a probléma célja - a túlélés - a kényelmes 0 vagy 1 formátumban van kifejezve.

John Wheeler az "It from Bit"-ben [7] azt állítja, hogy a világgegyetem bináris választásokon alapul. Hasonlóképpen, egy molekulákból álló emberek által működtetett vállalkozás valójában bináris bináris döntések sorozatára épül.

Ráadásul az adatok valós történelmi eseményen alapulnak, ami a mesterségesen létrehozott példákkal ellentétben a kutatás szempontjából értékessé teszi őket. Csak a Kaggle platformon, az egyik legnagyobb Data Pipeline és ETL, 1 355 998 ember vett részt a Titanic-adatkészleten alapuló kihívásokban, 53 963 egyedi Data Pipeline-megoldást fejlesztve [157] (9.2-9. ábra).

Hihetetlennek tűnik, de a Titanic utasairól szóló minden össze 1000 sornyi adat 12 paraméterrel hipotézisek, logikai láncok és egyedi adatvezetékek milliónak terepe lett. Egy kis adathalmazból végtelen számú meglátás, hipotézis és értelmezés születik - az egyszerű túlélési modellektől a rejtegett mintákat és a bonyolult érvelési labirintusokat figyelembe vevő összetett együttesekig.

Machine Learning from Disaster

[Submit Prediction](#)

Data Code Models Discussion Leaderboard Rules

Titanic Tutorial 16916 Gold ***

Updated 3y ago 29858 comments · Titanic - Machine Learning from Disaster

Titanic competition w/ TensorFlow Decision Forests 1098 Gold ***

Updated 2y ago Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

Titanic Data Science Solutions 10723 Gold ***

Updated 6y ago 2590 comments · Titanic - Machine Learning from Disaster

Exploring Survival on the Titanic 3968 Gold ***

Updated 7y ago Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster

Ábra 9.2-9 Az első öt megoldás az összesen 53 963 kész és nyílt forráskódú Pipeline megoldásból. Csak a Kaggle-on már közel 1,5 millióan próbálták megoldani ezt a feladatot [157].

Ha még egy ilyen kis táblázat is képes több millió egyedi megoldást generálni (9.2-9. ábra), mit mondhatunk a valós ipari építőipari adatsorokról, ahol a paraméterek tízezres nagyságrendben mérhetők?

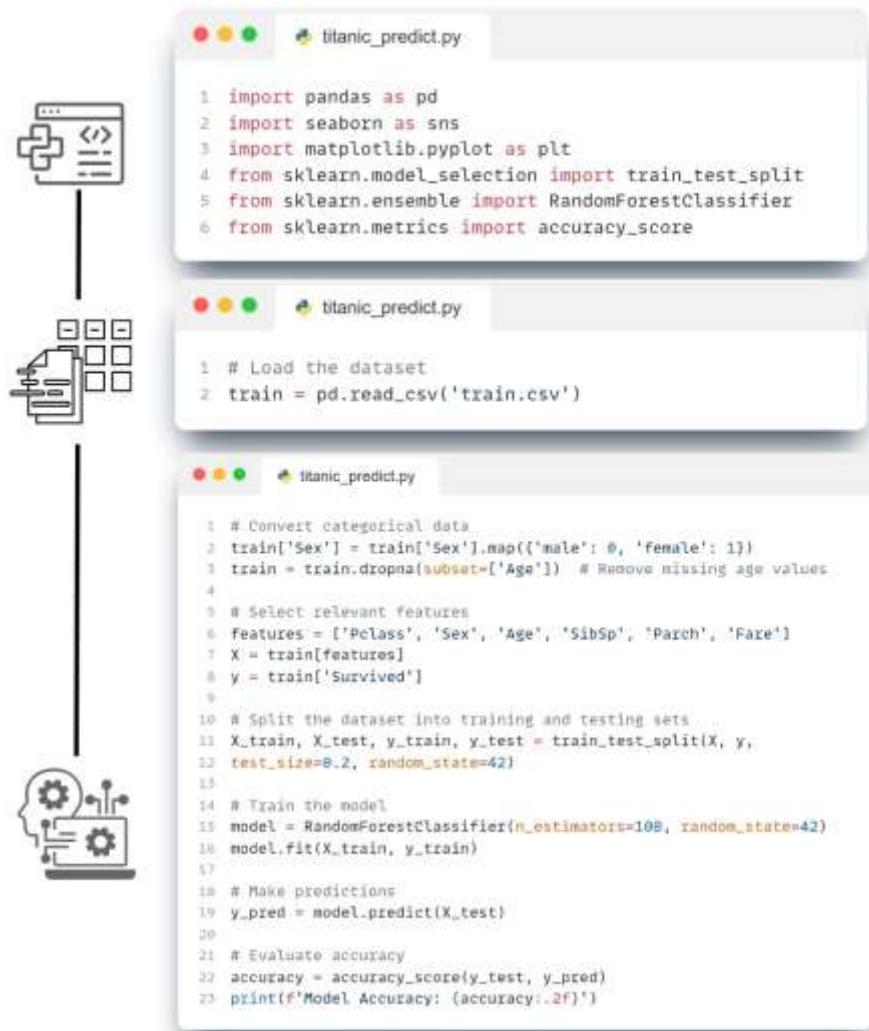
Egy viszonylag kis épület szabványos CAD -tervezése több tízezer entitást tartalmaz több ezer paraméterrel - a geometriai jellemzőktől a költség- és időattribútumokig. Képzelje el, hogy hány lehetséges meglátás, kapcsolat, előrejelzés és menedzsment-hipotézis rejtvőzik a vállalata összes projektének az évek során összegyűjtött adatai között. A történeti projektadatok nem csupán egy archívum - ezek egy szervezet élő emlékezete, annak digitális lábnyoma, amelyet elemezve számos egyedi hipotézis felállítása lehetséges.

A legfontosabb, hogy nem kell megvárnod, hogy a Kaggle közösség érdeklődjön a vállalatod vagy az adataid iránt. Elkezdhet dolgozni azzal, amivel ma rendelkezik: futtasson elemzést a saját adatain, képezzen modelleket a saját adatain, azonosítsa az ismétlődéseket, anomáliákat és a mintákat. Ahol korábban évekig tartó kísérletezésre és drága tanácsadásra volt szükség, ma már elég a kezdeményezés, egy LLM, az adatokhoz való nyitott hozzállás és a tanulási hajlandóság.

- ❸ A gépi tanulási algoritmus elkészítéséhez, amely a train.csv utasok képzési adatállománya alapján megjósolja az utasok túlélési arányát, kérjük meg az LLM -t, hogy oldja meg ezt a feladatot számunkra:

A Titanic utasainak képzési adathalmaza alapján készítsen egy gépi tanulási modellt a túlélési arányok előrejelzésére. ↴

2 LLM válasza :



```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

1 # Load the dataset
2 train = pd.read_csv('train.csv')

1 # Convert categorical data
2 train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
3 train = train.dropna(subset=['Age']) # Remove missing age values
4
5 # Select relevant features
6 features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
7 X = train[features]
8 y = train['Survived']
9
10 # Split the dataset into training and testing sets
11 X_train, X_test, y_train, y_test = train_test_split(X, y,
12 test_size=0.2, random_state=42)
13
14 # Train the model
15 model = RandomForestClassifier(n_estimators=100, random_state=42)
16 model.fit(X_train, y_train)
17
18 # Make predictions
19 y_pred = model.predict(X_test)
20
21 # Evaluate accuracy
22 accuracy = accuracy_score(y_test, y_pred)
23 print(f'Model Accuracy: {accuracy:.2f}')

```

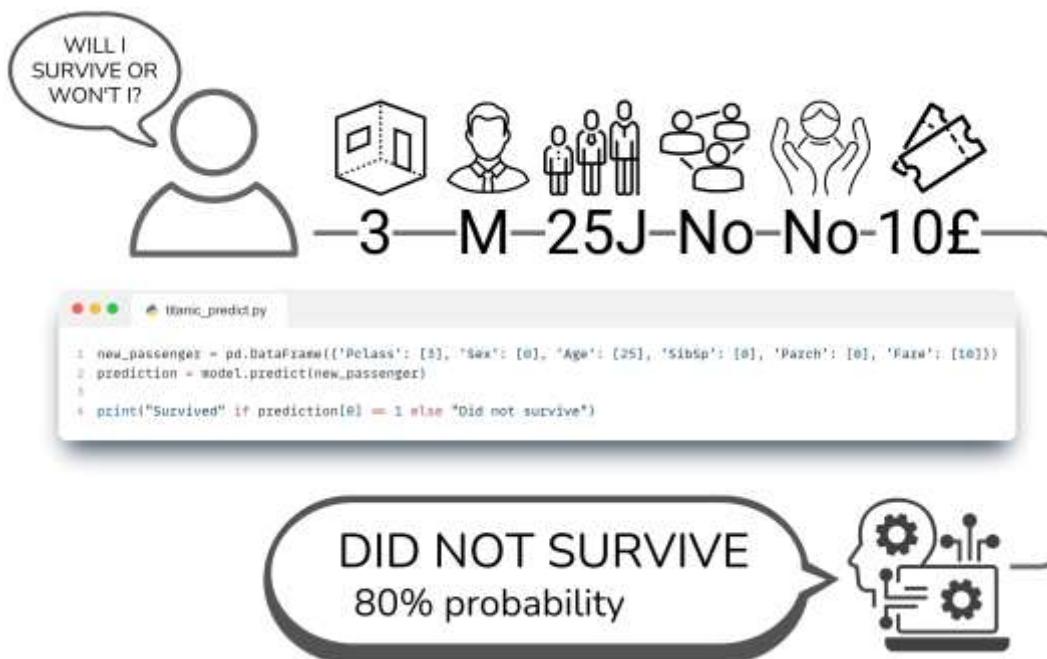
9.2-10 Az LLM Random Forest gépi tanulási algoritmus segítségével készített előrejelzést a Titanic túlélőiről.

Az LLM (9.2-10. ábra) eredményül kapott kódja betölti a Titanic utasainak adatait, megtisztítja azokat, a kategorikus változókat (pl. a nemet numerikus formátumba konvertálja), és a RandomForestClassifier algoritmus segítségével betanítja a modellt annak előrejelzésére, hogy egy utas túlélte-e vagy sem (a népszerű algoritmusokról a következő fejezetekben bővebben beszélünk).

A kód a képzési adatokat képzési és tesztkészletekre osztja (a Kaggle honlapján már kész test.csv (9.2-7. ábra) és train.csv (9.2-6. ábra) készleteket készített a képzéshez, majd a modellt a képzési adatokon képzik ki, és a tesztadatokon tesztelik, hogy megnézzék, mennyire jó egy adott előrejelző modell. A képzés után a test.csv-ből származó tesztadatokat (a túlélők és nem túlélők valós adataival) betapláljuk a modellbe, és az megjósolja, hogy ki élte túl és ki nem. A mi esetünkben a kapott gépi tanulási modell pontossága körülbelül 80%, ami azt mutatja, hogy elég jól megragadja a mintákat.

A gépi tanulás egy gyermekhez hasonlítható, aki egy téglalap alakú tömböt próbál beilleszteni egy kerek lyukba. A kezdeti szakaszokban az algoritmus számos megközelítést kipróbál, hibákkal és ellentmondásokkal találkozik. Ez a folyamat nem tűnik hatékonynak, de fontos tanulási lehetőséget biztosít: az egyes hibák elemzésével a modell javítja előrejelzéseit, és egyre pontosabb döntéseket hoz.

Most ez a modell (9.2-10. ábra) használható az új utasok túlélési arányának előrejelzésére, és például, ha a model.predict függvény segítségével utasinformációkkal tápláljuk meg a paraméterekkel: "férfi", "3. osztály", "25 éves", "nincs rokon a fedélzeten", a modell egy előrejelzést fog produkálni - azt, hogy az utas 80%-os valószínűséggel nem fogja túlálni a katasztrófát, ha 1912-ben a Titanicon volt (9.2-11. ábra).



9.2-11 ábra A fentiekben létrehozott modell 80%-os valószínűséggel képes megjósolni, hogy bármelyik új Titanic-utas túléli-e vagy sem.

A Titanic utasainak túlélési előrejelzési modellje egy sokkal tágabb koncepciót illusztrál: az építőiparban naponta több ezer szakember hoz hasonló "kettős" döntést - egy döntés, egy projekt, egy becslés, egy eszköz, nyereség vagy veszteség, biztonság vagy kockázat életéről vagy haláláról. Ahogy a Titanic példájában, ahol a kimenetet tényezőkön (nem, kor, osztály) múlott, az építőiparban a döntés minden egyéb szempontját számos saját tényező és változó (táblázatoszlopok) befolyásolja: anyagköltség, munkások képzettsége, határidők, időjárás, logisztika, műszaki kockázatok, megjegyzések és több százezer egyéb

paraméter.

Az építőiparban a gépi tanulás ugyanazokat az elveket követi, mint más területeken: a modelleket múltbeli adatokon - projektekből, szerződésekben, becslésekben - képzik ki, hogy teszteljék a különböző hipotéziseket és megtalálják a leghatékonyabb megoldásokat. Ez a folyamat olyan, mintha egy gyermeket próbáltatással és tévedéssel tanítanánk: minden egyes ciklusban a modellek alkalmazkodnak és egyre pontosabbá válnak.

A felhalmozott adatok felhasználása új távlatokat nyit az építés előtt. A munkaigényes kézi számítások helyett olyan modellek képezhetők ki, amelyek nagy pontossággal képesek megjósolni a jövőbeli projektek kulcsfontosságú jellemzőit. Ily módon a prediktív analitika az építőipart olyan térré alakítja át, ahol nemcsak tervezni, hanem magabiztosan előre jelezni is lehet a fejlesztéseket.

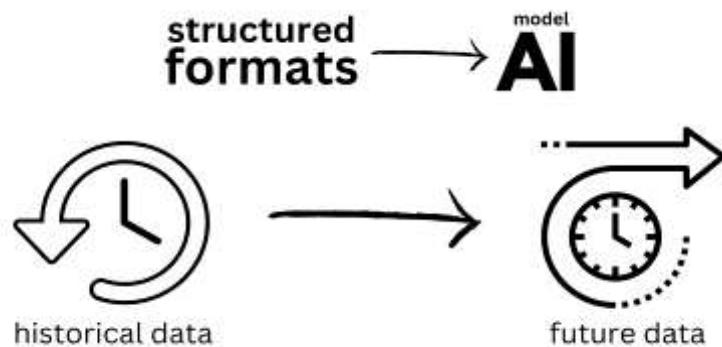
Előrejelzések és előrejelzések történeti adatok alapján

A vállalat projektjeiről gyűjtött adatok lehetővé teszik olyan modellek létrehozását, amelyek képesek a jövőbeni, még meg nem valósult objektumok költség- és időjellemzőinek előrejelzésére - időigényes kézi számítások és összehasonlítások nélkül. Ez lehetővé teszi az értékelési folyamatok jelentős felgyorsítását és egyszerűsítését, nem szubjektív feltételezésekre, hanem megalapozott matematikai előrejelzésekre támaszkodva.

Korábban, a könyv negyedik részében részletesen áttekintettük a projektköltségbecslés hagyományos módszereit, köztük az erőforrás-alapú módszert, valamint megemlíttetük a parametrikus és a szakértői megközelítéseket is. Ezek a módszerek még mindig relevánsak, de a modern gyakorlatban kezdik gazdagítani őket a statisztikai elemzés és a gépi tanulás eszközeivel, amelyek jelentősen javíthatják a becslések pontosságát és reprodukálhatóságát.

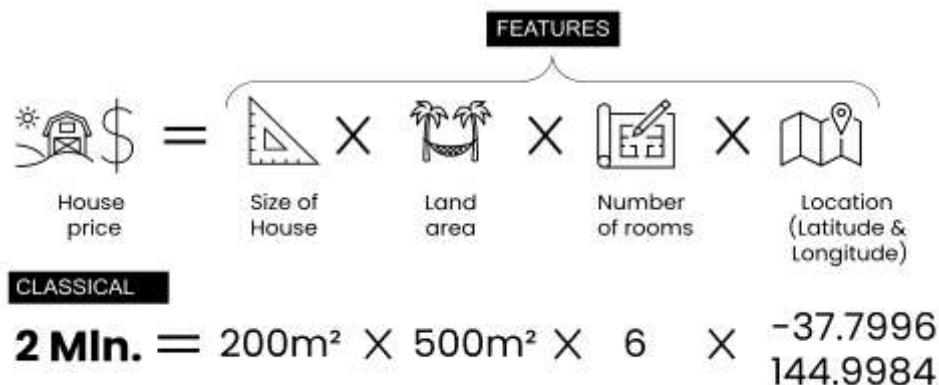
Az árak és időjellemzők kézi és félautomata kiszámításának folyamatait a jövőben kiegészítik a múltbeli adatok elemzésére, rejttet minták megtalálására és megalapozott döntések meghozatalára képes ML-modellek véleménye és előrejelzései. A már rendelkezésre álló információkból automatikusan új adatokat és forgatókönyveket generálnak majd - hasonlóan ahhoz, ahogyan a nyelvi modellek (LLM) az évek során nyílt forrásokból gyűjtött adatok alapján szövegeket, képeket és kódokat hoznak létre [158].

Ahogy ma az emberek a jövőbeli események értékelésében a tapasztalatra, az intuícióra és a belső statisztikára támaszkodnak, az elkövetkező években az építési projektek jövőjét egyre inkább a felhalmozott tudás és a matematikai gépi tanulási modellek kombinációja fogja meghatározni.



9.2-12 A minőségi és strukturált múltbeli vállalati adatok képezik azt az anyagot, amelyre a gépi tanulási modellek és előrejelzések épülnek.

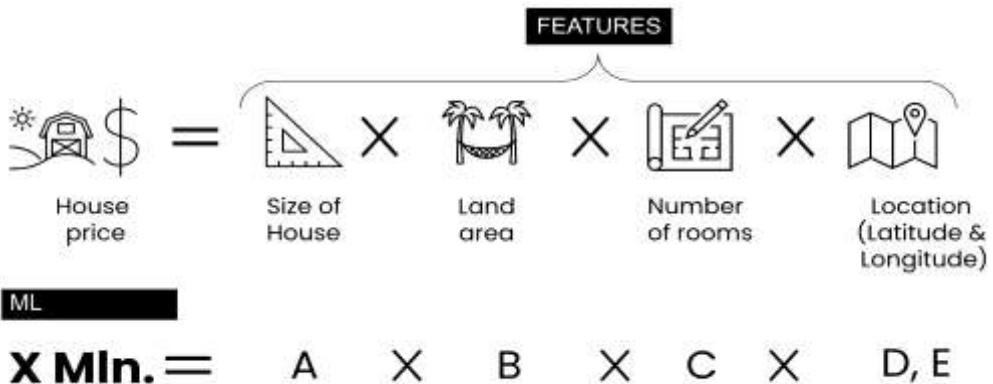
Vegyük egy egyszerű példát: egy ház árának előrejelzése a terület, a telek mérete, a szobák száma és a földrajzi elhelyezkedés alapján. Az egyik megközelítés egy klasszikus modell létrehozása, amely elemzi ezeket a paramétereket, és kiszámítja a várható árat (9.2-13. ábra). Ez a megközelítés pontos és előre ismert képletet igényel, ami a valós gyakorlatban gyakorlatilag lehetetlen.



Ábra 9.2-13 Egy klasszikus algoritmussal megbecsülhető egy ház értéke, amelynek egy rögzített képletét kell megtalálni.

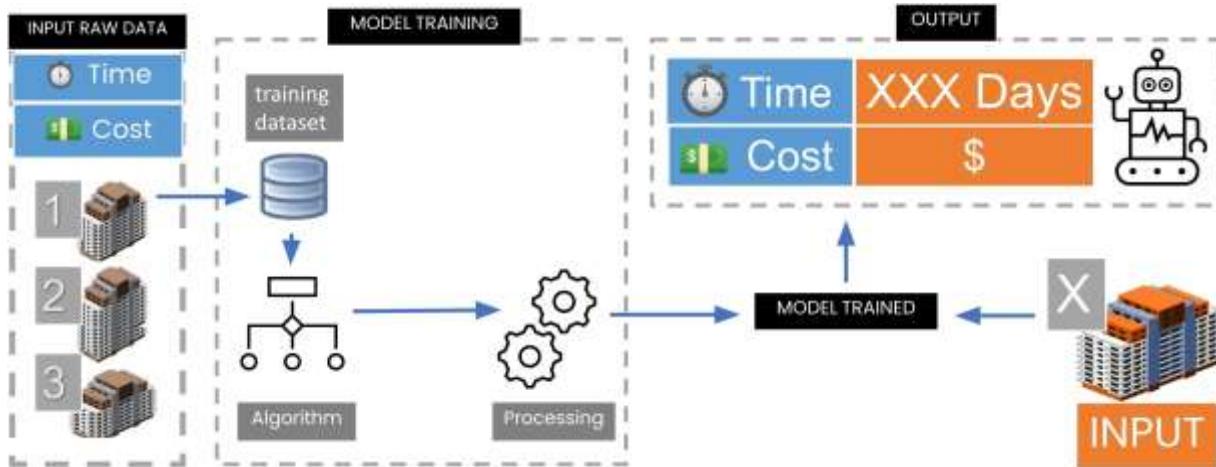
A gépi tanulás megszünteti a képletek kézi keresését, és képzett algoritmusokkal helyettesíti azokat, amelyek önállóan azonosítják a függőségeket, amelyek sokszor pontosabbak, mint bármely előre meghatározott egyenlet. Alternatív megoldásként hozunk létre egy gépi tanulási algoritmust, amely a probléma előzetes megértése és az esetlegesen hiányos múltbeli adatok alapján létrehoz egy modellt (9.2-14. ábra).

Az árképzés példáján keresztül a gépi tanulás lehetővé teszi különböző típusú matematikai modellek létrehozását, amelyek nem igénylik a költségképződés pontos mechanizmusának ismeretét. A modell a korábbi projektek adataiból "tanul", alkalmazkodva az építési paraméterek, azok költségei és a határidők közötti valós mintákhoz.



9.2-14 A klasszikus képletalapú becsléstől eltérően a gépi tanulási algoritmust a múltbeli adatokon képzik ki.

A felügyelt gépi tanulással összefüggésben, a képzési adathalmaz minden egyes projektje tartalmaz bemeneti attribútumokat (pl. hasonló épületek költsége és ideje) és várható kimeneti értékeket (pl. költség vagy idő). Egy hasonló adathalmazt használnak egy gépi tanulási modell létrehozásához és testreszabásához (9.2-15. ábra). Minél nagyobb az adathalmaz és minél jobb minőségűek az abban szereplő adatok, annál pontosabb lesz a modell és annál pontosabbak lesznek az előrejelzési eredmények.



9.2-15. ábra Egy korábbi projektek költség- és ütemezési adatain kiképzett ML-modell bizonyos valószínűséggel meghatározza egy új projekt költségét és ütemezését.

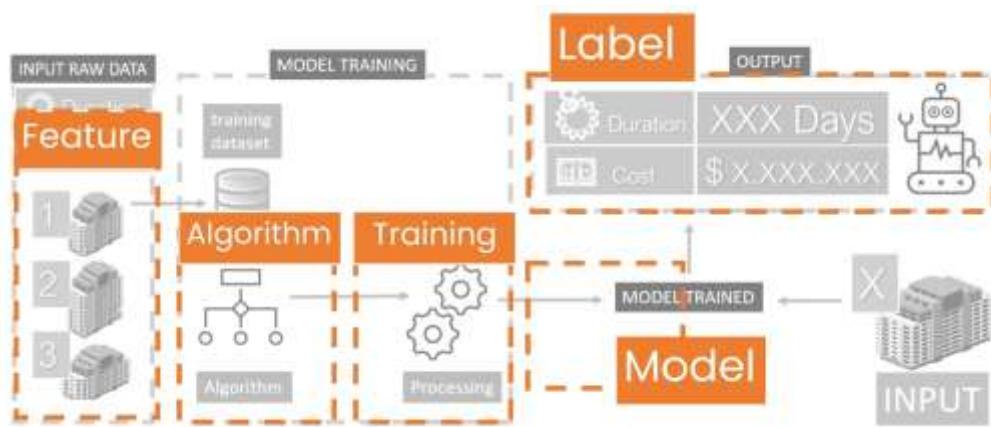
Miután a modell elkészült és betanították egy új projekt építésének megbecslésére, egyszerűen adja meg a modellnek az új projekt új attribútumait, és a modell a korábban megtanult minták alapján bizonyos valószínűséggel becsült eredményeket fog szolgáltatni.

A gépi tanulás kulcsfogalmai

A gépi tanulás nem varázslat, csak matematika, adatok és minták keresése. Nem rendelkezik valódi intelligenciával, hanem egy olyan program, amelyet adatokon képezz ki, hogy felismerje a mintákat és döntéseket hozzon állandó emberi beavatkozás nélkül.

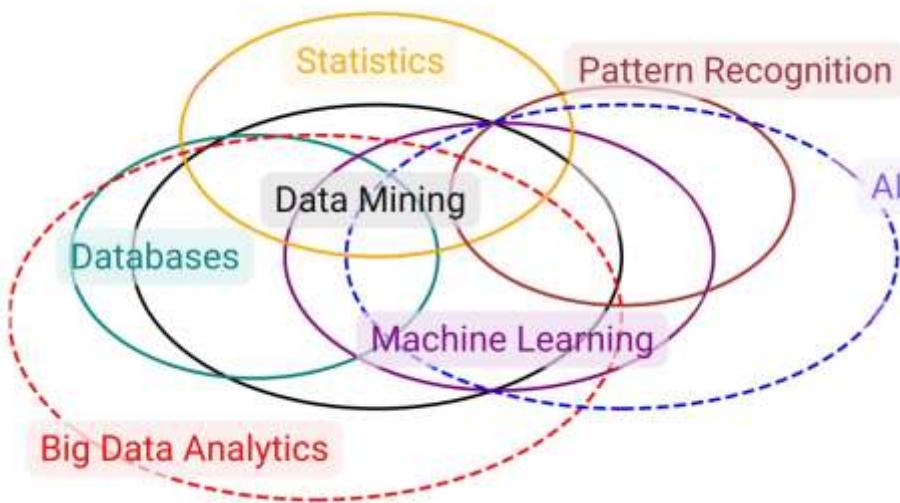
A gépi tanulás számos kulcsfogalmat használ szerkezetének leírására (9.2-16. ábra):

- **A címkek** olyan célváltozók vagy attribútumok (a Titanic-adatkészletben a "Survivor" paraméter), amelyeket a modellnek meg kell jósolnia. Példa: építési költség (pl. dollárban), az építési munkálatok időtartama (pl. hónapokban).
- **A jellemzők** olyan független változók vagy attribútumok, amelyek a modell bemenetként szolgálnak. Egy előrejelző modellben a címkek előrejelzésére szolgálnak. Példák: telekméret (négyzetméterben), az épület emeletszáma, az épület teljes alapterülete (négyzetméterben), földrajzi elhelyezkedés (földrajzi szélesség és hosszúság), az építés során felhasznált anyagok típusa. A jellemzők száma meghatározza az adatok dimenzióját.
- **modell** különböző hipotézisek halmaza, amelyek közül az egyik megközelíti a megjósolandó vagy közelítendő célfüggvényt. Példa: a gépi tanulási modell, amely regresszióelemzési technikákat használ az építkezések költségének és ütemezésének előrejelzésére.
- **Tanulási algoritmus** A **tanulási algoritmus** a legjobb hipotézis megtalálásának folyamata egy modellben, amely pontosan megfelel a célfüggvénynek egy képzési adathalmaz felhasználásával. Példa: lineáris regresszió, KNN vagy véletlen erdő algoritmus, amely a költség- és építési időadatokat elemzi az összefüggések és minták azonosítása érdekében.
- **Képzés** - A képzési folyamat során az algoritmus elemzi a képzési adatokat, és olyan mintákat talál, amelyek megfelelnek a bemeneti attribútumok és a célcímkek közötti kapcsolatnak. Ennek a folyamatnak az eredménye egy betanított gépi tanulási modell, amely készen áll az előrejelzésre. Példa: olyan folyamat, amelyben egy algoritmus elemzi a múltbeli építési adatokat (költség, idő, létesítményjellemzők), hogy előrejelző modellett hozzon létre.



9.2-16. ábra Az ML címkeket és attribútumokat használ olyan modellek létrehozásához, amelyeket algoritmusok segítségével képeznek ki az adatokon, hogy megjósolják az eredményeket.

A gépi tanulás nem elszigetelten létezik, hanem az analitikai tudományágak szélesebb ökoszisztemájának része, beleértve a statisztikát, az adatbázisokat, az adatbányászatot, a mintafelismerést, a nagy adatelemzést és a mesterséges intelligenciát. A 9.2-17. ábra bemutatja, hogy ezek a területek hogyan fedik és egészítik ki egymást, átfogó keretet biztosítva a modern döntéshozatali és automatizálási rendszerek számára.

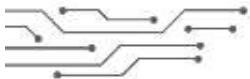


9.2-17. ábra Az adatelemzés különböző területei közötti kapcsolat: statisztika, gépi tanulás, mesterséges intelligencia, big data, mintafelismerés és adatbányászat.

A gépi tanulás fő célja, hogy a számítógépeket felruházza azzal a képességgel, hogy emberi beavatkozás vagy segítség nélkül, automatikusan tanuljanak tudást, és ennek megfelelően igazítsák a cselekvéseiket [159].

A jövőben tehát az ember szerepe csak az lesz, hogy kognitív képességekkel lássa el a gépet - ő fogja beállítani a feltételeket, súlyokat és paramétereket, a többöt pedig a gépi tanulási modell fogja elvégezni.

A következő fejezetben konkrét példákat fogunk megvizsgálni az algoritmusok alkalmazására. Valós táblázatok és egyszerűsített modellek segítségével mutatjuk be, hogyan épül fel lépésről lépésre az előrejelzés.



FEJEZET 9.3.

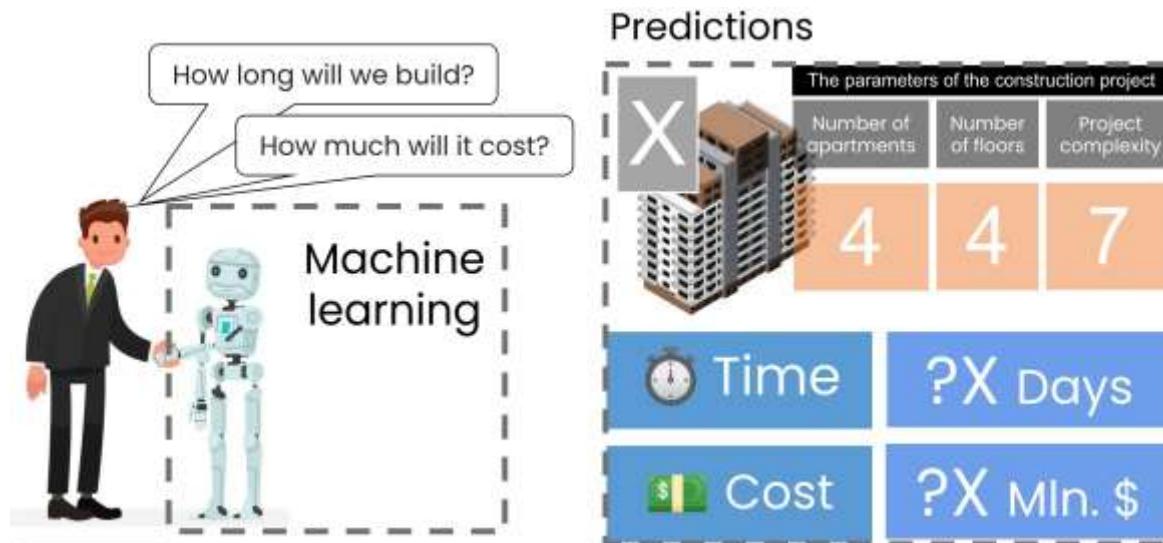
A KÖLTSÉGEK ÉS HATÁRIDŐK ELŐREJELZÉSE GÉPI TANULÁSSAL

Egy példa a gépi tanulás használatára a projektköltségek és határidők meghatározására.

Az építési idő és költségek becslése az egyik legfontosabb folyamat az építőipari vállalatok tevékenységében. Az ilyen becsléseket hagyományosan szakértők készítik a tapasztalatok, szakkönyvek és szabályozási adatbázisok alapján. A digitális átalakulással és a növekvő adatelérhetőséggel azonban ma már lehetőség van gépi tanulási (ML) modellek alkalmazására az ilyen becslések pontosságának javítása és automatizálása érdekében.

A gépi tanulás bevezetése az építési költségek és az építési ütemezés kiszámításának folyamatába nemcsak hatékonyabb tervezést tesz lehetővé, hanem kiindulópontot jelent az intelligens modellek más üzleti folyamatokba történő integrálásához is - a kockázatkezeléstől a logisztika és a beszerzés optimalizálásáig.

Fontos, hogy gyorsan meg tudja határozni, mennyi időbe telik egy projekt megépítése, és mennyi lesz a teljes költség. Ezek a projektidővel és költségekkel kapcsolatos kérdések hagyományosan az építőipar születése óta mind az ügyfelek, mind az építőipari cégek figyelmének előterében állnak.



9.3-1 Az építési projekteknél az építési idő és a költségek becslésének gyorsasága és minősége kulcsfontosságú sikertényező.

A következő példában a múltbeli projektekből származó kulcsadatokat kivonjuk és felhasználjuk a gépi tanulási modell kifejlesztéséhez, amely lehetővé teszi számunkra, hogy a modellt új paraméterekkel új építési projektek költségének és ütemezésének becslésére használjuk (9.3-1. ábra).

Tekintsünk három projektet három fő jellemzővel: a lakások száma (ahol a 100 lakás a könnyebb

szemléltetés érdekében a 10-es számnak felel meg), az emeletek száma és az építés bonyolultságának feltételes mérőszáma egy 1-től 10-ig terjedő skálán, ahol a 10 a legmagasabb bonyolultsági pontszám. A gépi tanulásban az olyan értékek, mint például a 100 10-re vagy az 50 5-re történő átalakításának és egyszerűsítésének folyamatát "normalizálásnak" nevezik.

A gépi tanulásban a **normalizálás** az a folyamat, amelynek során a különböző numerikus adatokat közös skálára hozzuk a feldolgozás és elemzés megkönnyítése érdekében. Ez a folyamat különösen fontos, ha az adatok különböző skálákkal és mértékegységekkel rendelkeznek.

Tegyük fel, hogy az első projekt (9.3-2. ábra) 50 lakással (normalizálás után 5), 7 emeettel és 2 komplexitási pontszámmal rendelkezik, ami viszonylag egyszerű konstrukciót jelent. A második projekt már 80 lakással, 9 emeettel és viszonylag összetett konstrukcióval rendelkezett. Ilyen feltételek mellett az első és a második lakóház építése 270, illetve 330 napot vett igénybe, a projekt teljes költsége pedig 4,5 millió, illetve 5,8 millió dollár volt.

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment s	Number of floors	Project complexity	Time	Cost
1	5	7	2	270	\$ 4.502.000
2	8	9	6	330	\$ 5.750.000
3	3	5	3	230	\$ 3.262.000
X	4	4	7	?X	\$?X. XXX.XXX

9.3-2. ábra Egy példa a múltbeli projektek egy csoportjára, amelyet a jövőbeli X projekt idejének és költségének becsléséhez használnak.

Amikor ilyen adatokra gépi tanulási modellt építünk, a fő feladat az előrejelzés szempontjából kritikus attribútumok (vagy címkék) azonosítása, jelen esetben az építési idő és a költségek. Egy kis adathalmazzal a korábbi építési projektekre vonatkozó információkat fogjuk felhasználni az új projektek tervezéséhez: a gépi tanulási algoritmusok segítségével egy új X projekt költségét és építési időtartamát kell megjósolnunk az új projekt adott attribútumai alapján, mint például 40 lakás, 4 emelet és egy viszonylag magas, 7-es projektkomplexitás (9.3-2. ábra). Valós körülmények között a bemeneti paraméterek száma sokkal nagyobb lehet, több tíz és több száz tényező között. Ezek közé tartozhatnak: az építőanyagok típusa, éghajlati övezet, a kivitelezők képzettségi szintje, a közművek elérhetősége, az alapozás típusa, a munkálatok megkezdésének évszaka, a művezetők megjegyzései stb.

Egy prediktív gépi tanulási modell létrehozásához, ki kell választanunk egy algoritmust. Egy algoritmus a gépi tanulásban olyan, mint egy matematikai recept, amely megtanítja a számítógépet arra, hogyan készítsen előrejelzéseket (a paraméterek megfelelő sorrendben történő keverése) vagy hozzon döntéseket az adatok alapján.

A múltbeli építési projektekre vonatkozó adatok elemzésére és a jövőbeli projektek idejének és költségeinek előrejelzésére (9.3-2. ábra) az egyik népszerű gépi tanulási algoritmus használható:

- **Lineáris regresszió (lineáris regresszió):** ez az algoritmus közvetlen kapcsolatot próbál találni az attribútumok között, például az emeletek száma és az építési költség között. Az algoritmus célja, hogy olyan lineáris egyenletet találjon, amely a legjobban leírja ezt a kapcsolatot, és amely lehetővé teszi előrejelzések készítését.
- **Algoritmus k-nearest neighbours (k-NN):** ez az algoritmus egy új projektet hasonlít össze korábbi, hasonló méretű vagy összetettségű projektekkal. A k-NN az alapján osztályozza az adatokat, hogy a k (számú) gyakorló példa közül melyik van hozzá a legközelebb. A regresszióval összefüggésben az eredmény a k legközelebbi szomszédok átlaga vagy mediánya.
- **Döntési fák:** olyan prediktív modellezési modell, amely az adatokat különböző feltételek alapján alcsoportokra osztja egy fa struktúra segítségével. A fa minden egyes csomópontja az adatok további felosztásához vezető feltételt vagy kérdést, és minden egyes levél a végső előrejelzést vagy eredményt jelöli. Az algoritmus az adatokat kisebb csoportokra osztja különböző jellemzők alapján, például először a történetek száma, majd a komplexitás és így tovább, hogy előrejelzést készítsen.

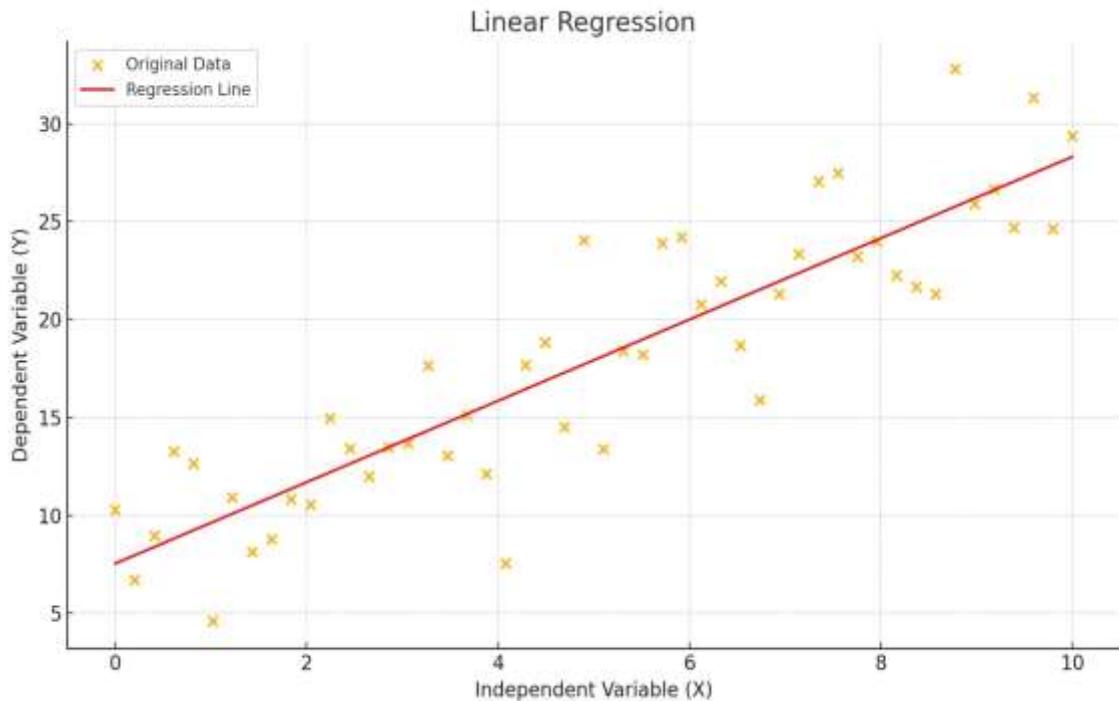
Nézzük meg a gépi tanulási algoritmusokat egy új projekt költségének becslésére két népszerű algoritmus példáján: a lineáris regresszió és a K-közeli szomszédok algoritmus.

Projektköltség és -idő előrejelzése lineáris regresszió segítségével

A lineáris regresszió egy alapvető adataelemzési algoritmus, amely egy változó értékét egy vagy több másik változóval való lineáris kapcsolat alapján jósolja meg. Ez a modell feltételezi, hogy a függő változó és egy vagy több független változó között közvetlen lineáris kapcsolat van, és az algoritmus célja ennek a kapcsolatnak a megtalálása.

A lineáris regresszió egyszerűsége és áttekinthetősége miatt számos területen népszerű eszközzé vált. Ha egyetlen változóval foglalkozunk, a lineáris regresszió lényege, hogy megtaláljuk a legjobban illeszkedő egyenest az adatpontokon keresztül.

A lineáris regresszió megtalálja a legjobb egyenest (piros vonal), amely közelíti a függősséget az X bemeneti változó és az Y kimeneti változó között. Ez az egyenes lehetővé teszi az Y értékek előrejelzését új X értékekre az azonosított lineáris kapcsolat alapján (9.3-3. ábra).



Ábra 9.3-3 A lineáris regresszió elve az, hogy a legjobb egyenest keressük a képzési értékeken keresztül.

Ezt a vonalat egy egyenlet ábrázolja, amelyben a független változó (X) értékének megadása a függő változó (Y) előre jelzett értékét eredményezi. Ez a folyamat lehetővé teszi, hogy az Y hatékonyan megjósolható legyen az X ismert értékeiből a köztük lévő lineáris kapcsolat segítségével. Egy ilyen statisztikailag átlagolt egyenes megtalálására példa a San Francisco-i építési engedélyezési adatok értékelése (9.1-7. ábra), ahol a lineáris regresszió segítségével számították ki az inflációt a különböző típusú létesítményekre.

Töltsük be a projekt adattáblázatát (9.3-2 ábra az előző fejezetből) közvetlenül az LLM programba, és kérjük meg, hogy készítsen nekünk egy egyszerű gépi tanulási modellt.

✉ Küldjön szöveges kérést az LLM chat (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Meg kell mutatnunk egy egyszerű gépi tanulási modell felépítését, amely egy új X projekt megvalósításának költségeit és idejét jelzi előre (9.3-2 ábra a mellékelt képen). ↴

■ LLM válasza :

The screenshot shows a software interface for AI development. At the top, there's a toolbar with icons for 'Create in LLM Chat' and four circular icons representing different AI models or functions. Below the toolbar is a code editor window titled 'Linear regression.py'. The code itself is as follows:

```

1 from sklearn.linear_model import LinearRegression
2 import numpy as np
3
4 # Data from the image
5 # Features: Number of apartments, number of floors, project complexity
6 X = np.array([
7     [5, 7, 2],
8     [8, 9, 6],
9     [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

At the bottom of the code editor, there's a 'Run in IDE' button. Below the code editor, a 'Result' section displays the output of the code execution:

```
(array([238.44444444]), array([3042337.77777778]))
```

A 9.3-4 ábra LLM a lineáris regressziót választotta a gépi tanulási modell létrehozásához a projekt költségének és idejének előrejelzésére.

Az LLM automatikusan felismerte a táblázatot a csatolt képből, és az adatokat vizuális formátumból táblázatos tömbre konvertálta (9.3-4. ábra - 6. sor). Ez a tömb szolgált alapul a jellemzők és címkék létrehozásához, amelyekből egy gépi tanulási modell készült (9.3-4. ábra - 17-22. sor), amely lineáris regressziót használt.

Egy "rendkívül kis" adathalmazon betanított lineáris regressziós modell segítségével előrejelzéseket készítettek egy új hipotetikus építési projektre vonatkozóan, amelyet X projektnek neveztek el. A problémánkban ezt a projektet 40 lakás, 4 emelet és 7-es komplexitási szint jellemzi (9.3-2. ábra).

Ahogyan azt az új X projektre vonatkozó korlátozott és kis adathalmazon alapuló lineáris regressziós modell előre jelezte (9.3-4. ábra - 24-29. sor):

- **Az építkezés** időtartama körülbelül 238 nap (238.444444444)
- **A teljes kiadás** körülbelül \$ 3,042,338 (3042337.777) lesz.

A projektköltség-hipotézis további vizsgálatához hasznos a különböző gépi tanulási algoritmusokkal és módszerekkel kísérletezni. Ezért jósoljuk meg ugyanazokat a költség- és időértékeket egy új X projektre egy kis mennyiségű múltbeli adat alapján a K-Nearest Neighbours algoritmus (k-NN) segítségével.

A projektköltségek és -idők előrejelzése a K-közelibbi szomszéd algoritmus segítségével (k-NN)

A k-Nearest Neighbours (k-NN) algoritmust használjuk további előrejelzőként egy új projekt költségének és időtartamának becsléséhez. A K-Nearest Neighbours (k-NN) algoritmus egy felügyelt gépi tanulási (felügyelt gépi tanulási) módszer mind az osztályozásra, mind a regresszióra. A k-NN algoritmust korábban már tárgyaltuk a vektoradatbázisok keresésével összefüggésben is (8.2-2. ábra), ahol a legközelebbi vektorok (pl. szövegek, képek vagy műszaki leírások) megtalálására használják. Ebben a megközelítésben minden projekt egy többdimenziós tér egy pontjaként jelenik meg, ahol minden dimenzió a projekt egy adott attribútumának felel meg.

A mi esetünkben az egyes projektek három attribútumát egy háromdimenziós tér pontjaiként fogjuk ábrázolni (9.3-5. ábra). Így a közelgő X projektünket ebben a térben (x=4, y=4, z=7) koordinátákkal fogjuk lokalizálni. Megjegyzendő, hogy valós körülmények között a pontok száma és a tér dimenzionalitása nagyságrendekkel nagyobb lehet.

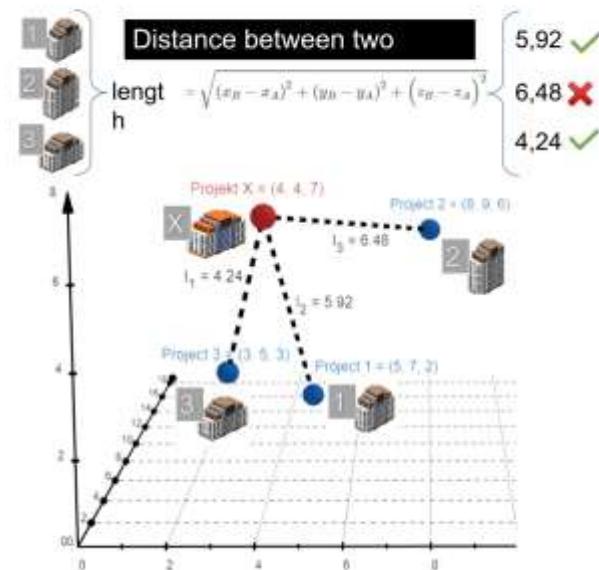
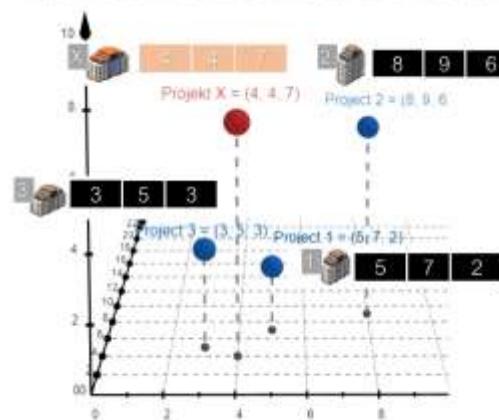
A K-NN (k-nearest neighbours) algoritmus a kívánt X projekt és a képzési adatbázisban található projektek közötti távolság mérésével működik. E távolságok összehasonlításával az algoritmus meghatározza azokat a projekteket, amelyek a legközelebb vannak az új X projekt pontjához.

Például, ha az eredeti adatállományunk második projektje (x=8, y=9, z=6) sokkal távolabb van X-től (9.3-5. ábra), mint a többi projekt, akkor kizártató a további elemzésből. Ennek eredményeként csak a két (k=2) legközelebbi projektet lehet felhasználni a számításokhoz, amelyek alapján az átlagértéket határozzuk meg.

Egy ilyen módszer a szomszédsági keresés révén lehetővé teszi a projektek közötti hasonlóságok értékelését, ami viszont segít következtetéseket levonni egy új projekt valószínűsíthető költségeire és időzítésére vonatkozóan a korábban megvalósított hasonló projektek alapján.

k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



Ábra 9.3-5 A K-NN algoritmusban a projektek egy többdimenziós térben pontokként vannak ábrázolva, és a legközelebbi projektek a távolságok alapján kerülnek kiválasztásra a hasonlóság értékeléséhez és előrejelzéshez.

A k-NN munkája több kulcsfontosságú lépést foglal magában:

- **Adatfelkészítés:** először a gyakorló és tesztadatkészleteket töltjük be. A képzési adatok az algoritmus "betanítására" szolgálnak, a tesztadatok pedig az algoritmus hatékonyságának ellenőrzésére.
- **A K paraméter kiválasztása:** kiválasztunk egy K számot, amely megadja, hogy hány legközelebbi szomszédot (adatpontot) kell figyelembe venni az algoritmusban. A "K" értéke nagyon fontos, mert befolyásolja az eredményt.
- Osztályozási folyamat és regresszió a tesztadatokra:
 - **Távolságok kiszámítása:** a tesztadatok minden egyes elemére kiszámítjuk a távolságot a képzési adatok minden egyes elemétől (9.3-5. ábra). Ehhez különböző távolságmérési módszerek használhatók, például az euklideszi távolság (a legelterjedtebb módszer), a Manhattan-távolság vagy a Hamming-távolság.
 - **Rendezés és K legközelebbi szomszéd kiválasztása:** a távolságok kiszámítása után a távolságok rendezése és a vizsgálati ponthoz legközelebbi K pont kiválasztása.
 - **A vizsgálati pont osztályának vagy értékének meghatározása:** ha osztályozási feladatról van szó, akkor a vizsgálati pont osztályát a K kiválasztott szomszédok közül a leggyakoribb osztály alapján határozzuk meg. Ha regressziós feladatról van szó, akkor a K szomszéd értékének átlagát (vagy más központi tendencia mérőszámát) kell kiszámítani.
- **A folyamat befejezése:** ha az összes vizsgálati adatot besorolták, vagy előrejelzéseket készítettek rájuk vonatkozóan, a folyamat befejeződött.

A k-nearest neighbours (k-NN) algoritmus számos gyakorlati alkalmazásban hatékony, és az egyik legfontosabb eszköz a gépi tanulással foglalkozó szakemberek eszköztárában. Ez az algoritmus egyszerűsége és hatékonysága miatt népszerű, különösen olyan feladatokban, ahol az adatok közötti kapcsolatok könnyen értelmezhetők.

Példánkban a K-közelibbi szomszédok algoritmus alkalmazása után azonosítottuk azt a két projektet (a kis mintánkból), amelyek a legrövidebb távolságra vannak az X projekt től (9.3-5. ábra). Ezen projektek alapján az algoritmus meghatározza az áruk és az építési időtartamuk átlagát. Az elemzést követően (9.3-6. ábra) az algoritmus a legközelebbi szomszédok átlagolásával arra a következtetésre jut, hogy az X projekt körülbelül \$ 3 800 000 \$-ba fog kerülni, és körülbelül 250 napot vesz igénybe a kivitelezés.

k-nearest neighbors algorithm



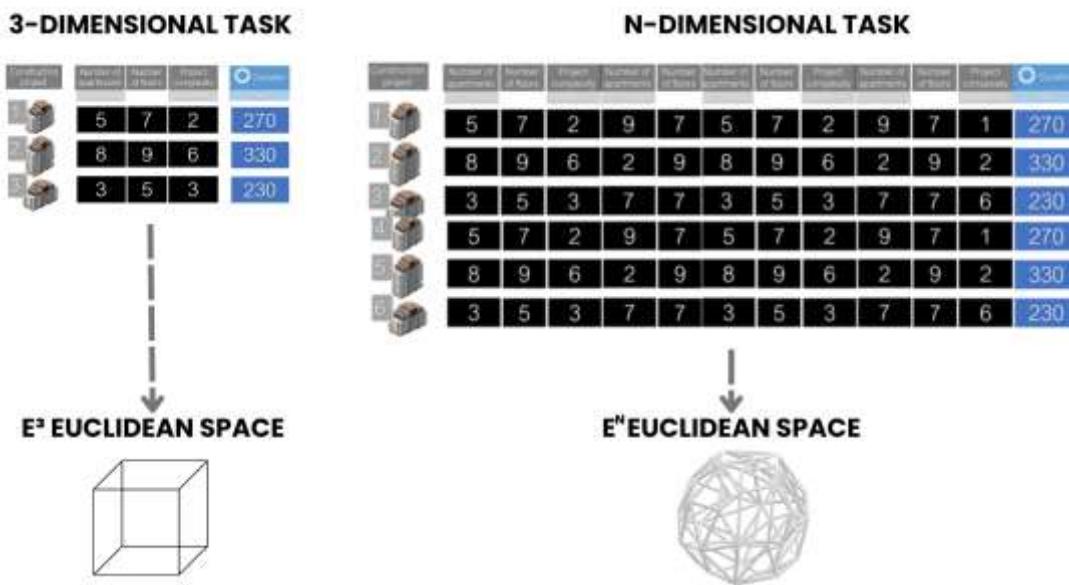
9.3-6 ábra A K-közelibbi szomszédok algoritmus a mintában lévő két legközelebbi projekt elemzésével határozza meg az X projekt költségeit és ütemezését.

A k-Nearest Neighbors (k-NN) algoritmus különösen népszerű a osztályozási és regressziós feladatokban, például az ajánlórendszerben, ahol az adott felhasználó érdeklődési köréhez hasonló preferenciák alapján javasol termékeket vagy tartalmakat. Ezenkívül a k-NN-t széles körben használják az orvosi diagnosztikában a betegségtípusok osztályozására a beteg tünetei alapján, a mintafelismerésben és a pénzügyi szektorban az ügyfelek hitelképességének értékelésére.

A gépi tanulási modellek még korlátozott adatok esetén is hasznos előrejelzéseket adhatnak, és jelentősen javíthatják az építési projektmenedzsment analitikus komponensét. A múltbeli adatok bővítésével és tisztításával lehetőség nyílik a kifinomultabb modellek felé való elmozdulásra - például az építkezés típusának, helyének, az építkezés megkezdésének évszakának és egyéb tényezőknek a figyelembevételével.

Egyszerűsített feladatunk három attribútumot használt a 3D térben történő megjelenítéshez, de a valós

projektek átlagosan több száz vagy ezer attribútumot tartalmaznak (lásd az "Egy példa a CAD-alapú nagy adatokra (BIM)" című fejezetben található adatállományt), ami jelentősen növeli a tér dimenzionalitását és a projektek vektorként való ábrázolásának összetettségét (9.3-7. ábra).



9.3-7 ábra Az egyszerűsített példában három attribútumot használtunk a 3D -visualizációhoz, míg a valós projekteknél ennél több van.

A különböző algoritmusok alkalmazása ugyanarra az X projektre vonatkozó adathalmazra, amely 40 lakással, 4 emelettel és 7-es komplexitási szinttel rendelkezik, különböző előrejelzett értékeket eredményezett. A lineáris regressziós algoritmus 238 napos befejezési időt és 3 042 338\$ költséget jósolt (9.3-4. ábra), míg a k-NN algoritmus 250 napot és 3 882 000\$ költséget jósolt (9.3-6. ábra).

A gépi tanulási modellek segítségével kapott előrejelzések pontossága közvetlenül a bemeneti adatok mennyiségtől és minőségtől függ. Minél több projektet vonnak be a képzésbe, és minél teljesebb és pontosabb a jellemzők (attribútumok) és eredményeik (címkek) reprezentációja, annál nagyobb a valószínűsége annak, hogy megbízható, minimális hibaértékkal rendelkező előrejelzéseket kapunk.

A adatfelidolgozási technikák fontos szerepet játszanak ebben a folyamatban, többek között:

- Normalizálás a jellemzők közös skálára való hozatalához;
- Kiugró értékek felismerése és kiküszöbölése, a modell torzításának kiküszöbölése;
- Kategorikus attribútumok kódolása a szöveges adatok manipulálásának lehetővé tétele érdekében;
- Hiányzó értékek kitöltése, a modell robusztusságának növelése.

Ezenkívül keresztpotenciális módszereket alkalmaznak a modell általánosíthatóságának és új adathalmazokra való robusztusságának értékelésére a túllesztés felderítése és az előrejelzés megbízhatóságának javítása érdekében.

A káosz egy megfejthető rend [160].

- José Saramago, "A kettős"

Még ha úgy is tűnik Önnek, hogy az Ön feladatainak káoszát nem lehet formálisan leírni, tudnia kell, hogy a világ bármely eseménye és különösen az építési folyamatok matematikai törvényszerűségeknek vannak alávetve, amelyekhez szükség lehet az értékek kiszámításának támogatására nem szigorú képletekkel, hanem statisztikák és történelmi adatok segítségével.

Mind a hagyományos költségszámítási osztály becslései, mind a gépi tanulási modellek elkerülhetetlenül bizonytalansággal és potenciális hibaforrásokkal szembesülnek. Ha azonban elegendő minőségi adat áll rendelkezésre, a gépi tanulási modellek hasonló, sőt néha még a szakértői becsléseknel is nagyobb előrejelzési pontosságot tudnak felmutatni.

A gépi tanulás valószínűleg megbízható kiegészítő eszközzé válik az elemzésekhez, amely képes: finomítani a számításokat, alternatív forgatókönyveket javasolni, és azonosítani a projektparaméterek közötti rejtett függvégeket. Az ilyen modellek nem fogják azt állítani, hogy univerzálisak, de hamarosan fontos helyet foglalnak majd el a számítási és döntéshozatali folyamatokban. A gépi tanulási technológiák nem fogják kizárnia a mérnökök, becslők és elemzők részvételét, hanem éppen ellenkezőleg, a múltbeli adatokon alapuló további nézőpontot kínálva bővíti képességeiket.

Ha megfelelően integráljuk az építőipari vállalatok üzleti folyamataiba, a gépi tanulás a vezetői döntéstámogató rendszer fontos elemévé válhat - nem az emberek helyettesítésére, hanem szakmai intuíciójuk és mérnöki logikájuk kiterjesztéseként.

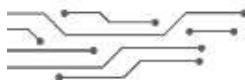
Következő lépések: a tárolástól az elemzésig és előrejelzésig

Az adatokkal való munka modern megközelítései kezdik megváltoztatni a döntéshozatal módját az építőiparban. Az intuitív értékelésekről az objektív adatelemzésre való áttérés nemcsak a pontosságot javítja, hanem új lehetőségeket is nyit a folyamatok optimalizálására. E rész összegzéseként érdemes kiemelni a főbb gyakorlati lépéseket, amelyek segítségével a tárgyalt módszereket alkalmazhatja minden nap során:

- Fenntartható tárolási infrastruktúra létrehozása
 - Próbálja meg a különböző dokumentumokat és projektadatokat egyetlen táblázatos modellben egyesíteni, a kulcsfontosságú információkat egyetlen adatkeretben összesíteni a további elemzéshez.
 - Használjon hatékony adattárolási formátumokat - például oszlopos formátumokat, mint az Apache Parquet a CSV vagy XLSX helyett - különösen azoknál a halmazoknál, amelyeket a jövőben gépi tanulási modellek képzésére lehet használni.
 - Adatverzió-rendszer létrehozása a változások nyomon követése érdekében a projekt során.
- Elemzési és automatizálási eszközök bevezetése

- Kezdje el a korábbi projektadatok elemzését - dokumentáció, modellek, becslések alapján - a minták, trendek és anomáliák azonosítása érdekében.
 - ETL-folyamatok (Extract, Transform, Load) automatikus betöltése és előkészítése az adatok automatikus betöltése és előkészítése érdekében.
 - Tanulja meg, hogyan lehet a legfontosabb mérőszámokat különböző ingyenes Python vizualizációs könyvtárak segítségével megjeleníteni.
 - statisztikai módszerek és véletlenszerű mintavételezés alkalmazásának megkezdése reprezentatív és reprodukálható analitikai eredmények előállítása érdekében.
- Az adatokkal való munka egyre érettebbé válása
- Néhány alapvető gépi tanulási algoritmus elsajátítása egyszerű és egyszerű példákon, mint például a Titanic adathalmazon.
 - Elemezze a jelenlegi folyamatokat, és állapítsa meg, hogy hol lehet a merev ok-okozati logikától az előrejelzés és becslés statisztikai módszerei felé elmozdulni.
 - Kezdje el az adatokat stratégiai eszközökkel kezelni, nem pedig melléktermékként: a döntéshozatali folyamatokat inkább az adatmodellek, mint a konkrét szoftvermegoldások köré építse.

Azok az építőipari vállalatok, amelyek felismerték az adatok értékét, a fejlődés új szakaszába lépnek, ahol a versenyelőnyt nem az erőforrások mennyisége, hanem az analitikán alapuló döntéshozatal sebessége határozza meg.



MAXIMÁLIS KÉNYELEM A NYOMTATOTT VERZIÓVAL

Ön birtokában van a **Data-Driven Construction** ingyenes digitális verziója. A kényelmesebb munkavégzés és az anyagokhoz való gyors hozzáférés érdekében javasoljuk, hogy figyeljen a nyomtatott kiadásra:



Mindig kéznél van: a nyomtatott formátumú könyv megbízható munkaeszközévé válik, amely lehetővé teszi, hogy bármilyen munkaszituációban gyorsan megtalálja és használja a megfelelő ábrákat és diagramokat.

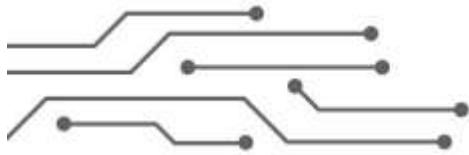
Kiváló minőségű illusztrációk: a nyomtatott kiadásban szereplő összes kép és grafika maximális minőségen jelenik meg.

Gyors hozzáférés az információkhoz: könnyű navigáció, jegyzetek, könyvjelzők készítése és a könyvvel való munka bárhol.

A könyv teljes nyomtatott változatának megvásárlásával egy kényelmes eszközt kap az információkkal való kényelmes és hatékony munkavégzéshez: a képi anyagok gyors használatának lehetőségét a minden nap feladatakban, a szükséges sémkák gyors megtalálását és jegyzetelését. Emellett vásárlásával támogatja a nyílt tudás terjesztését.

Rendelje meg a könyv nyomtatott változatát a következő címen:

datadrivenconstruction.io/books



X RÉSZ

AZ ÉPÍTŐIPAR A DIGITÁLIS KORBAN. LEHETŐSÉGEK ÉS KIHÍVÁSOK

Az utolsó, tizedik rész átfogóan áttekinti az építőipar jövőjét a digitális átalakulás korában. Elemzi az ok-okozati elemzéstől a nagy adatkorrelációkkal való munkavégzésre való áttérést. Párhuzamot von a képzőművészeti fejlődése és az építőipari adatmunka fejlődése között, bemutatva, hogy az ágazat a részletes ellenőrzéstől a folyamatok holisztkus megértése felé halad. Megvizsgálják az építőipar "uberizációjának" koncepcióját, ahol az adatok átláthatósága és az automatizált számítások radikálisan megváltoztathatják a hagyományos üzleti modelleket, megszüntetve a közvetítők szükségességét és csökkentve a spekulációs lehetőségeket. A megoldatlan kérdéseket, mint például a elemek egyetemes osztályozása, részletesen tárgyalják, időt adva az építőipari vállalatoknak az új környezethez való alkalmazkodásra. A rész konkrét ajánlásokkal zárul a digitális átalakulási stratégia kialakítására, amely magában foglalja a sebezhetősségek elemzését és a szolgáltatások bővítését, hogy versenyképesek maradhassanak a változó iparágban.

FEJEZET 10.1.

TÚLÉLÉSI STRATÉGIÁK: VERSENYELŐNYÖK KIÉPÍTÉSE

Korrelációk helyett számítások: az építőipari analitika jövője

Az információ gyors digitalizálódásának köszönhetően (1.1-5. ábra) a modern építőipar alapvető átalakuláson megy keresztül, ahol az adatok nem csupán eszközök, hanem olyan stratégiai eszközök válnak, amelyek alapvetően megváltoztathatja a projekt- és üzletvezetés hagyományos megközelítéseit.

Az építési tevékenységek évezredek óta determinisztikus módszerekre - pontos számításokra, részletezésre és a paraméterek szigorú ellenőrzésére - támaszkodnak. Az i. sz. első évszázadokban a római mérnökök matematikai elveket alkalmaztak a vízvezetékek és hidak építésénél. A középkorban az építészek a gótikus katedrálisok ideális arányaira törekedtek, a 20. századi iparosodás során pedig szabványosított normák és előírások rendszerei alakultak ki, amelyek a tömegépítés alapjává váltak.

Napjainkban a fejlődés vektora a szigorúan ok-okozati összefüggések keresésétől a valószínűségi elemzés, a korrelációk és rejtett minták keresése felé tolódik el. Az iparág új szakaszba lép - az adatok kulcsfontosságú erőforrássá válnak, és az ezeken alapuló analitika felváltja az intuitív és helyileg optimalizált megközelítéseket.



Ábra 10.1-1 Az építési adatokban rejlik rejtett potenciál: a vállalatnál meglévő számítások csak a jéghegy csúcsát jelentik, amely a vezetőség számára elemezhető.

Egy vállalat információs rendszere olyan, mint egy jéghegy (10.1-1. ábra): az adatok potenciáljának csak egy kis része látható a vállalat vezetése számára, míg a fő érték a mélyben rejtőzik. Fontos, hogy az adatokat ne csak a jelenlegi felhasználásuk alapján értékeljük, hanem a jövőben felszabaduló lehetőségek alapján is. Azok a vállalatok lesznek azok, amelyek megtanulják, hogyan lehet az adatokból rejtett mintákat kinyerni és új felismeréseket létrehozni, amelyek fenntartható versenyelőnyre tesznek szert

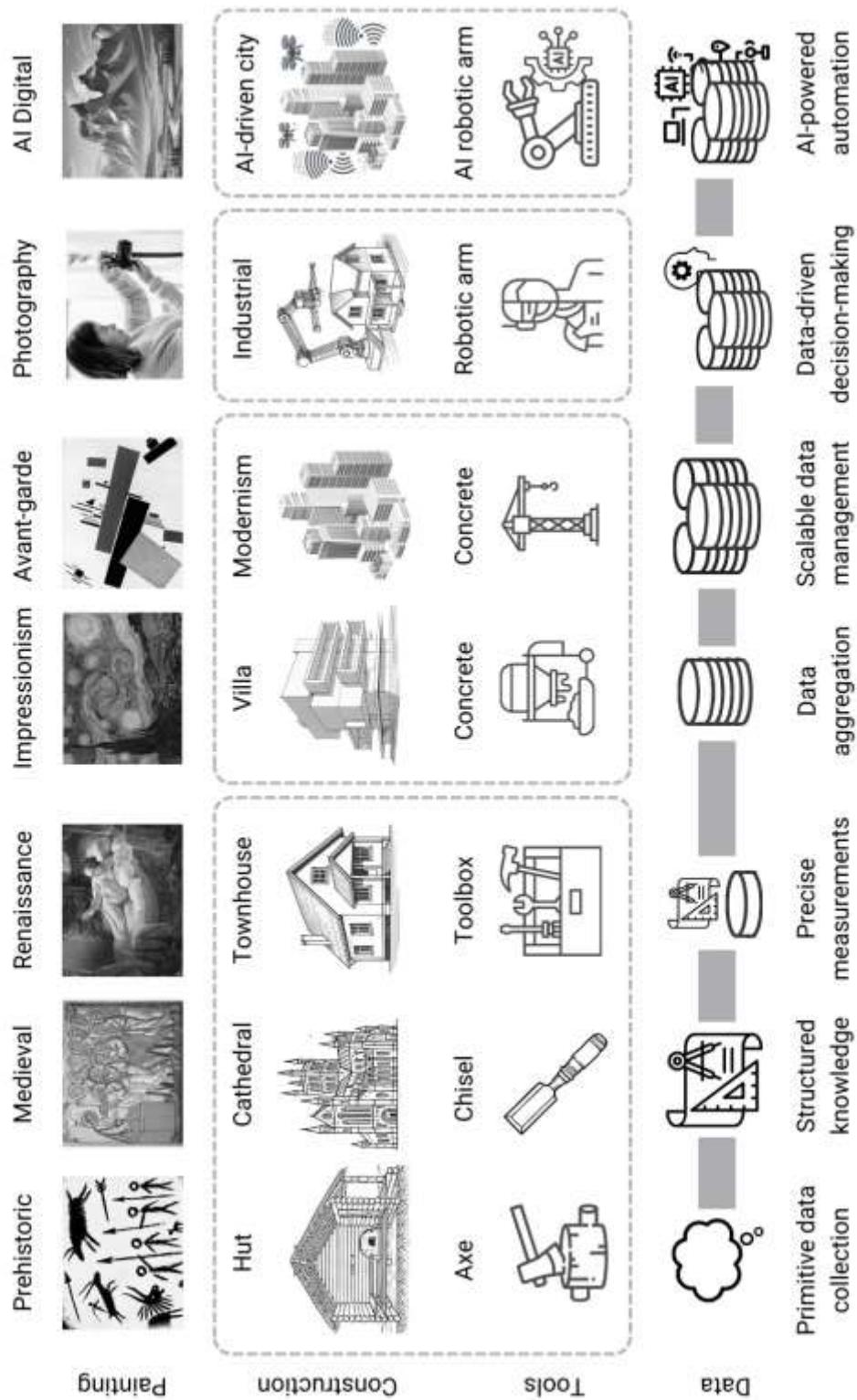
A rejtett minták megtalálása és az adatok értelmezése nem csupán a számokkal való munkáról szól, hanem egy olyan kreatív folyamatról, amely elvont gondolkodást és a széttagolt elemek mögötti teljes kép meglátásának képességét igényli. Ebben az értelemben az adatokkal való munka fejlődése a vizuális művészletek fejlődéséhez hasonlítható (10.1-2. ábra).

Az építőipar fejlődése figyelemre méltóan hasonlít a képzőművészeti fejlődéséhez. Az emberiség mindenkorban a kezdetleges módszerektől jutott el a kifinomult megjelenítési és elemzési technológiáig. A történelem előtti időkben az emberek barlangrajzokat és kezdetleges eszközöket használtak a minden nap feladataik megoldására. A középkor és a reneszánsz során az építészet és a művészet kifinomultsági szintje drámaian megnőtt. A kora középkorra az építőszerszámok az egyszerű fejszétől a technikai tudás növekedését szimbolizáló kiterjedt szerszámkészletekké fejlődtek.

A realizmus kora volt az első forradalom a képzőművészetben: a művészek megtanulták, hogyan kell a legapróbb részleteket is reprodukálni, a maximális hitelesség elérése érdekében. Az építőiparban ennek a korszaknak a megfelelője a precíz mérnöki technikák, a részletes rajzok és a szigorúan szabályozott számítások voltak, amelyek évszázadokra a tervezési gyakorlat alapjává váltak.

Később az impresszionizmus megváltoztatta a művészi valóság felfogását: a formák szó szerinti megjelenítése helyett a művészek a hangulatot, a fényt és a dinamikát kezdték el megragadni, és az abszolút pontosság helyett inkább egy átfogó benyomás visszatükrözésére törekedtek. Hasonlóképpen, a gépi tanulás az épületelemzésben a merev logikai modellektről a mintafelismerés és a valószínűségi minták felé mozdul el, amelyek lehetővé teszik, hogy "meeglássuk" az adatokban a klasszikus elemzés során hozzáférhetetlen rejtett függőségeket. Ez a megközelítés a Bauhaus minimalizmus és funkcionális eszméit idézi, ahol a jelentés (funkció) fontosabb, mint a forma. A Bauhaus a felesleges dolgok eltávolítására, a díszítés elhagyására törekedett a letisztultság, a használhatóság és a tömegvonzó jelleg kedvéért. A dolgoknak érhetőnek és hasznosnak kellett lenniük, felesleg nélkül - az esztétika a tervezés és a cél logikából született.

A 19. század végén a fotográfia megjelenésével a művészet új eszközzel gazdagodott, amellyel a valóságot soha nem látott pontossággal lehetett megörökíteni, és amely felforgatta a képzőművészethez való hozzáállást. Hasonlóképpen, az építőiparban a 21. századi ipari forradalom a robottechnológiák, lézerek, IoT, RFID és az olyan koncepciók, mint a Connected Construction, ahol az egyes paraméterek gyűjtése az építkezés teljes valóságának skálázható intelligens rögzítésévé fejlődött.



A 10.1-2. ábra Era of Fine Art Evolutions összhangban van az építőiparban az adatokkal való munkavégzés megközelítéseiben fejlődéssel.

Ma, ahogy a képzőművészet a mesterséges intelligencia eszközeinek és az LLM megjelenésével újragondoláson megy keresztül, az építőipar egy újabb kvantumugrást él át: a mesterséges intelligencia által vezérelt intelligens rendszerek (AI), az LLM-chats lehetővé teszik az előrejelzést, az optimalizálást és a megoldások generálását minimális emberi beavatkozással.

Az adatok szerepe a tervezésben és az irányításban gyökeresen megváltozott. Míg korábban az ismereteket szóban és empirikus jelleggel adták át - ahogyan a valóságot a 19. századig kézzel festett képekkel örökölték meg -, ma az építési "kép" teljes digitális rögzítése áll a középpontban. A gépi tanulási algoritmusok segítségével ez a digitális kép az építési valóság impresszionistikus ábrázolásává alakul át - nem pontos másolattá, hanem a folyamatok általánosított, valószínűségi megértésévé.

Gyorsan közeledünk egy olyan korszakhoz, amelyben az épületek tervezési, építési és üzemeltetési folyamatait nem csak kiegészítik, hanem nagyrészt mesterséges intelligencia rendszerek fogják irányítani. Ahogyan a modern digitális művészet ecset nélkül - szöveges utasítások és generatív modellek segítségével - jön létre, úgy a jövő építészeti és mérnöki megoldásait is a felhasználó által beállított kulcsfontosságú kérdések és paraméterek fogják alakítani.

A 21. században az adatokhoz való hozzáférés, azok értelmezése és az elemzések minősége nélkülözhetetlenné válik a projektek sikeréhez. Az adatok értékét pedig nem a mennyiségeük határozza meg, hanem az, hogy a szakemberek képesek-e elemezni, ellenőrizni és cselekvésre váltani azokat.

Adatvezérelt megközelítés az építőiparban: az infrastruktúra új szintje

Az emberiség történetében minden ilyen technológiai ugrás alapvető változásokat hozott a gazdaságban és a társadalomban. Napjainkban a 19. századi ipari forradalomhoz hasonló mértékű átalakulás új hullámának vagyunk tanúi. Míg azonban száz évvel ezelőtt a változás fő mozgatórugója a mechanikus erők és az energiatechnológiák voltak, addig most az adatok és a mesterséges intelligencia.

A gépi tanulás, az LLM és az AI ágensek megváltoztatják az alkalmazások lényegét, szükségtelenné téve a hagyományos szoftvercsomagokat (amelyeket a könyv második részében tárgyalunk) (2.2-3. ábra). minden adatlogika az AI-ügynökökben összpontosul, nem pedig a keményen kódolt üzleti szabályokban (2.2-4. ábra).

Az adatok korában az alkalmazások hagyományos szemlélete alapvetően átalakul. Egy olyan modell felé haladunk, amelyben a terjedelmes, moduláris vállalati rendszerek elkerülhetetlenül átadják a helyüket a nyílt, könnyű, testre szabott megoldásoknak.

A jövőben csak a mögöttes adatstruktúra marad meg, és minden interakció az adatbázissal közvetlenül együttműködő ügynökökön keresztül történik majd. Tényleg úgy vélem, hogy az egész alkalmazáshalmaz el fog tűnni, mert egyszerűen nincs rá szükség, amikor a mesterséges intelligencia közvetlenül a mögöttes adatbázissal lép kölcsönhatásba. Az egész karrieremet a SaaS - cégek építésével, azokban való munkával töltöttem, és hogy őszinte legyek, valószínűleg most nem indítanék új SaaS-vállalkozást. És valószínűleg most nem fektetnék be SaaS-cégekbe. A helyzet túl bizonytalan. Ez nem azt jelenti, hogy a jövőben nem lesznek szoftvercégek, csak azt, hogy nagyon másképp fognak kinézni. A jövő rendszerei több adattárral fognak dolgozni, nem korlátozódnak egyetlen adatbázisra. Minden logika a mesterséges intelligencia rétegbe fog költözni [46].

- Matthew Berman, a Forward Future vezérigazgatója

Az új paradiigma legfontosabb különbsége a technológiai ballaszt minimalizálása. A monumentálisan összetett és zárt szoftverrendszerek helyett rugalmas, nyitott és gyorsan testre szabható modulokat kapunk, amelyek szó szerint "élnek" az adatáramlásban (7.4-1. ábra - Apache Airflow, NiFi). A jövő folyamatmenedzsmentjének architektúrája mikroalkalmazások használatát irányozza elő - kompakt, célzott eszközök, amelyek alapvetően különböznek a masszív és zárt ERP, PMIS, CDE, CAFM rendszerektől. Az új ügynökök a lehető legadaptívabbak, integráltak és üzlet-specifikusak lesznek (pl. Low-Code/No-Code, 7.4-6. ábra).

Az összes üzleti logika ezekhez az [AI] ügynökökhöz kerül, és ezek az ügynökök CRUD [Create, Read, Update, and Delete] műveleteket hajtanak végre több tárolón, ami azt jelenti, hogy nem tesznek különbséget a használt háttértár között. Több adatbázist fognak frissíteni, és az összes logika az úgynevezett AI-szintre kerül. És amint az AI-szint lesz az, ahol az összes logika van, az emberek elkezdik majd lecserélni a backendeket. Már most is elég nagy százalékban látjuk a Dynamics backendek és az ügynökhelyszíni piaci győzelmét, és agresszívan fogunk ebbe az irányba haladni, megpróbáljuk mindezt összehozni. Legyen szó akár az ügyfélszolgálatról, akár más területekről, például nem csak a CRM, hanem a pénzügyi és üzemeltetési megoldásainkról is. Mert az emberek több olyan AI-vezérelt üzleti alkalmazást szeretnének, ahol a logikai réteget AI és AI-ügynökök vezérelhetik. [...]. Az egyik legizgalmasabb dolog számomra az Excel a Python, ami hasonló a GitHubhoz a Copilot. Vagyis amit mi csináltunk: most, hogy van Excel, csak nyissa meg, futtassa a Copilotot, és kezdjen el játszani vele. Ez már nem csak a rendelkezésre álló számokat érti - magától fog tervet készíteni. Ahogy a GitHub Copilot Workspace létrehoz egy tervet, majd végrehajtja azt, olyan ez, mint egy adatelemző, aki az Excel-t eszközökön használja a sorok és oszlopok vizualizálására az elemzéshez. A Copilot tehát az Excel-t eszközökön használja, annak minden képességevel együtt, mivel képes adatokat generálni, és Python-interpreterrel rendelkezik.

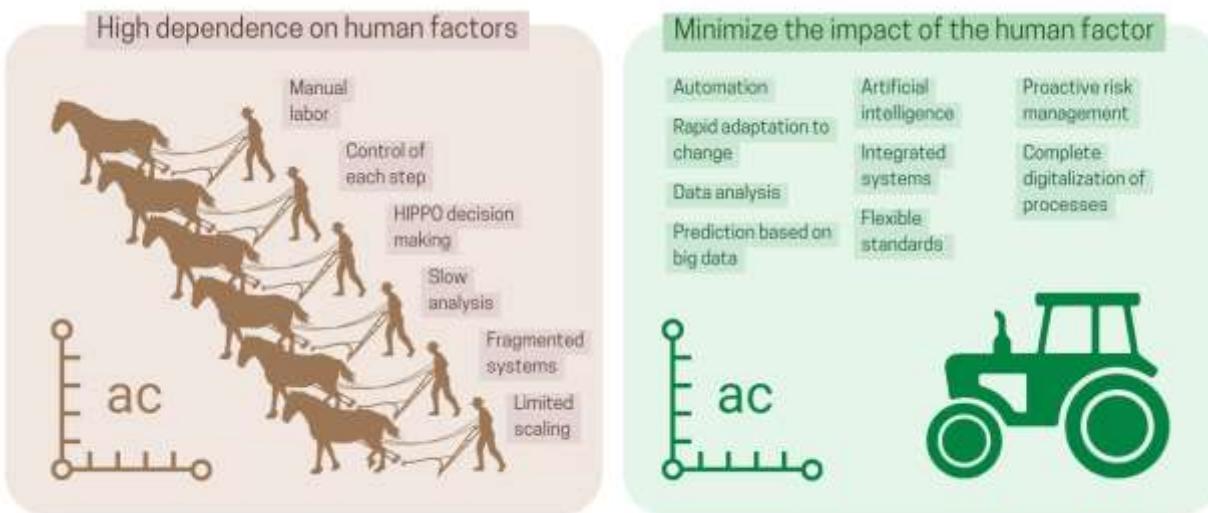
- Satya Nadella, vezérigazgató, Microsoft, interjú a BG2 csatornával 2024. december. [28]

Az irodai alkalmazások logikájában megfigyelhető átalakulás - a moduláris, zárt rendszerekről a nyílt adatokkal közvetlenül együttműködő mesterséges intelligencia-ügynökök felé történő elmozdulás - csak

egy sokkal nagyobb folyamat része. Ez nem csupán az interfészkek vagy a szoftverarchitektúra megváltoztatásáról szól: a változások a munkaszervezés, a döntéshozatal és az üzleti irányítás alapelveit érintik. Az építőiparban ez egy olyan adatvezérelt logikához vezet, amelyben az adatok lesznek a folyamatok középpontjában a tervezéstől az erőforrás-menedzsmentig és az építésfelügyeletig.

A következő generációs digitális iroda: hogyan változtatja meg az AI a munkaterületet?

Közeli egy évszázaddal ezelőtt az emberiség már hasonló technológiai forradalmat élt át. A gőzgépekről az elektromos motorokra való áttérés több mint négy évtizedig tartott, de végül példátlan termelékenységnövekedést katalizált - elsősorban az energiakapacitás decentralizálása és az új megoldások rugalmassága révén. Ez a váltás nemcsak a történelem menetét változtatta meg, mivel a népesség nagy részét a vidéki területekről a városokba költözött, hanem a modern gazdaság alapjait is lefektette. A technológia története a fizikai munkától az automatizálás és az intelligens rendszerek felé vezető út. Ahogyan a traktor felváltotta a földet művelő munkások sokaságát, úgy váltja fel a modern digitális technológia a hagyományos irodai építésirányítási módszereket (10.1-3. ábra). Még a 20. század elején a világ lakosságának nagy része kézzel művelte a földet, amíg az 1930-as években meg nem kezdődött a munka gépesítésének gépekkel és traktorokkal történő gépesítése.



Ábra 10.1-3 Ahogy a 20. század elején a traktor emberek tucatjait váltotta fel, a 21. században a gépi tanulás fogja felváltani a hagyományos üzleti és projektmenedzsment módszereket.

Ahogyan az emberiség száz évvel ezelőtt az egyes földterületek kezdetleges eszközökkel történő megművelésétől a gépek segítségével történő nagyüzemi gazdálkodásig jutott el, úgy ma az információk szétszórta "silóinak" feldolgozásától az adatok tömbjeinek feldolgozásáig jutunk el a nagy teljesítményű "traktorok" - ETL-csatorna és mesterséges intelligencia algoritmusok - segítségével.

Egy hasonló ugrás küszöbén állunk - de digitális síkon: a hagyományos, manuális üzletmenedzsmenttől az adatvezérelt modellek felé.

A teljes körű adatvezérelt architektúrához vezető út időt, befektetést és szervezeti erőfeszítéseket igényel.

Ez az út azonban nem csak fokozatos javulás, hanem minőségi ugrás felé, az építési folyamatok nagyobb hatékonyisége, átláthatósága és irányíthatósága felé nyitja meg az utat. Mindez a digitális eszközök szisztematikus bevezetésétől és az elavult üzleti gyakorlatok elhagyásától függ.

A feladatparaméterezés, az ETL, az LLM, az IoT-összetevők, az RFID, a tokenizáció, a big **data** és a gépi tanulás **adatvezérelt építéssé** alakítja át a hagyományos építőipart, ahol a projekt és az építőipari tevékenység minden részletét az adatok fogják ellenőrizni és optimalizálni.

Korábban több ezer munkaórát vett igénybe az információk elemzése. Ma már ezeket a feladatokat algoritmusok és LLM-ek végzik, amelyek a széttagolt adathalmazokból stratégiai forrásokat alakítanak ki súgók segítségével. A technológiai világban ugyanaz történik, ami a mezőgazdaságban: a kapától az automatizált agrárgazdaság felé haladunk. Így az építőiparban is az irodai munka - az Excel-fájloktól és a kézi összegzéstől - egy intelligens rendszer felé mozdul el, ahol az adatokat összegyűjtik, megtisztítják, strukturálják és meglátásokká alakítják.

A vállalatoknak már ma el kell kezdeniük az információs mezők "megművelését" minőségi adatgyűjtéssel és az információk strukturálásával, majd tisztító és normalizáló eszközökkel kell "megtermékenyíteniük" őket, hogy aztán prediktív analitika és automatizált megoldások formájában "betakarítsák" őket. Ha egy modern gazda egy géppel képes helyettesíteni száz talajművelőt, az intelligens algoritmusok képesek lesznek levenni a rutint az alkalmazottakról, és az információáramlás stratégiai irányítónak szerepébe helyezni őket.

Fontos azonban megérteni, hogy egy valóban adatvezérelt -szervezet létrehozása nem egy gyors folyamat. Ez egy hosszú távú stratégiai irány, hasonlóan ahhoz, mintha egy új területet hoznánk létre a rendszerek új erdejének (1.2-5. ábra) termesztéséhez, ahol ebben az ökoszisztemámában minden egyes "fa" egy különálló folyamat, kompetencia vagy eszköz, amelynek növekedése és fejlődése időt vesz igénybe. Ahogy egy valódi erdő esetében, a sikер nemcsak az ültetési anyag minőségétől (technológia), hanem a talajtól (vállalati kultúra), az éghajlattól (üzleti környezet) és a gondozástól (rendszerszemléletű megközelítés) is függ.

A vállalatok többé már nem támaszkodhatnak kizárálag zárt, kész megoldásokra. A technológiai fejlődés korábbi szakaszaitól eltérően a jelenlegi átmenetet - a nyílt adathozzáférés, a mesterséges intelligencia és a nyílt forráskód felé - a nagy gyártók valószínűleg nem fogják támogatni, mivel az közvetlenül veszélyeztető a bevett üzleti modelljüket és az alapvető bevételi forrásaiat.

A Harvard Business School tanulmánya [40] szerint, amelyet már a negyedik és ötödik technológiai forradalomról szóló fejezetben is tárgyalunk, a leggyakrabban használt nyílt forráskódú megoldások nulláról történő létrehozása az összes vállalat számára körülbelül 4,15 milliárd dollárba kerülne. Ha azonban úgy képzeljük el, hogy minden vállalat saját alternatívákat fejlesztene ki a meglévő nyílt forráskódú eszközökhöz való hozzáférés nélkül, ami az elmúlt évtizedekben meg is történt, akkor a teljes üzleti költség elérheti a kolossal 8,8 millió dollárt - ez a az irracionális kereslet ára, amivel a szoftverpiacot értékelni lehet.

A technológiai fejlődés elkerülhetetlenül a bevett üzleti modellek újragondolásához vezet. Míg korábban a vállalatok komplex, átláthatatlan folyamatokból és zárt adatokból tudtak pénzt csinálni, az AI és az analitika fejlődésével ez a megközelítés egyre kevésbé életképes.

Az adatokhoz és eszközökhöz való hozzáférés demokratizálódása következtében a hagyományos szoftverértékesítési piac jelentősen zsugorodhat. Ezzel egyidejűleg azonban egy új piac fog növekedni - a digitális szakértelem, a testreszabás, az integráció és a megoldástervezés piaca. Itt az érték nem a licencértékesítésből, hanem a rugalmas, nyitott és alkalmazkodó digitális folyamatok kialakításának képességeből származik majd. Ahogyan a villamosítás és a traktorok megjelenése új iparágakat hozott létre, úgy a big data, a mesterséges intelligencia és az LLM alkalmazása is teljesen új távlatokat nyit az építőiparban, ami nemcsak technológiai beruházásokat, hanem a gondolkodásmód, a folyamatok és a szervezeti struktúrák mélyreható átalakítását is megköveteli. És azok a vállalatok és szakemberek, akik ezt felismerik és már ma elkezdenek cselekedni, a holnap vezetői lesznek.

Egy olyan világban, ahol a nyílt adatok egyre nagyobb értékké válnak, az információk hozzáférhetősége megváltoztatja a játékszabályokat. A befektetők, az ügyfelek és a szabályozó hatóságok egyre inkább átláthatóságot fognak követelni, és a gépi tanulási algoritmusok képesek lesznek automatikusan azonosítani az eltéréseket a becslésekben, határidőkben és költségekben. Ez megteremti a digitális átalakulás új szakaszának feltételeit, amely fokozatosan az építőipar "uberizációjához" vezet.

A nyílt adatok és az uberizáció veszélyt jelent a meglévő építőipari vállalkozásokra nézve

Az építőipar információkezelési folyamattá válik. Minél pontosabbak, minőségibbek és teljesebbek az adatok, annál hatékonyabb az épületek tervezése, számítása, költségbecslése, felépítése és üzemeltetése. A jövőben a legfontosabb erőforrás nem a daru, a beton és a betonacél lesz, hanem az információgyűjtés, -elemzés és -felhasználás képessége.

Az építőipari vállalatok ügyfelei - befektetők és az építkezést finanszírozó ügyfelek - a jövőben elkerülhetetlenül hasznosítani fogják a nyílt adatok és az analitika értékét a múltbeli adatokból. Ez lehetőségeket nyit majd a projekt ütemterveinek és költségeinek automatizált kiszámítására anélkül, hogy az építőipari vállalatokat költségszámítási kérdésekbe kellene bevonni, ami segít a költségek ellenőrzésében és a felesleges költségek gyorsabb azonosításában.

Képzeljünk el egy építkezést, ahol lézerszkennerek, kvadrokopterek és fotogrammetriai rendszerek gyűjtenek pontos, valós idejű adatokat a felhasznált beton mennyiségéről. Ezeket az információkat automatikusan egyszerű, sík MESH -modellé alakítják át metaadatokkal, megkerülve a nehézkes CAD (BIM) rendszereket, a bonyolult geometriai kernelek, ERP vagy PMIS függősége nélkül. Ezeket az építkezésről gyűjtött adatokat központilag egyetlen strukturált adattárba továbbítják, amely az ügyfél számára független elemzés céljából elérhető, és ahová különböző építőipari üzletek valós árait töltik fel, a hitelfinanszírozás mértékétől kezdve a dinamikusan változó tényezőig terjedő paramétereikkal, mint például az időjárási viszonyok, az építőanyagok tőzsdei jegyzései, a logisztikai tarifák és a munkaerőárak statisztikai szezonális ingadozásai. Ilyen körülmények között a tervezési és a tényleges anyagmennyiségek közötti eltérések azonnal nyilvánvalóvá válnak, így a becslések manipulálása mind a tervezési szakaszban, mind a projekt átadásakor lehetetlenné válik. Ennek eredményeképpen az építési folyamat átláthatósága nem felügyelők

és menedzserek seregével, hanem objektív digitális adatokkal valósul meg, ami minimálisra csökkenti az emberi tényezőt és a spekuláció lehetőségét.

A jövőben ezt a fajta adatellenőrzési munkát inkább az ügyféloldalon lévő adatkezelők fogják végezni (1.2-4. ábra CQMS manager). Ez különösen igaz a kalkulációkra és a projektbecslésekre: ahol korábban egy egész becslőosztály működött, ott holnap már gépi tanulással és előrejelző eszközökkel fognak rendelkezni, amelyek árkorlátokat állítanak fel az építőipari cégek számára, amelyekbe be kell illeszkedniük.

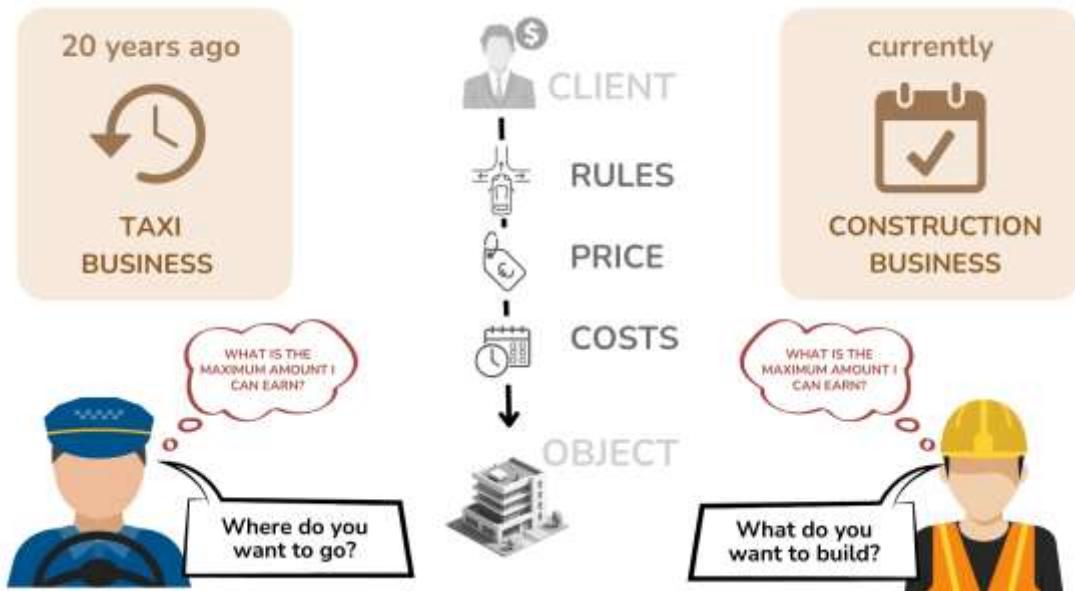
Tekintettel az [építőipari] ágazat széttöredezett jellegére, ahol a rendszerek és alrendszerök többségét kkv-k szállítják, a digitális stratégiának az ügyféltől kell kiindulnia. Az ügyfeleknek kell megteremteniük a feltételeket és mechanizmusokat az ellátási lánc digitális képességeinek felszabadításához [20].

- Andrew Davis és Giuliano Denicol, Accenture "Nagyobb értékteremtés a tőkeprojektek révén"

Az adatok ilyen mértékű nyitottsága és átláthatósága veszélyt jelent az építőipari vállalatokra, amelyek hozzászoktak ahhoz, hogy átláthatatlan folyamatokból és zavaros jelentésekből keressenek pénzt, ahol a spekulációk és a rejtett költségek bonyolult és zárt formátumok és moduláris, saját adatplatformok mögött rejtőzhetnek. Ezért az építőipari vállalatok - a nyílt forráskódú megoldásokat népszerűsítő szállítókhöz hasonlóan - valószínűleg nem érdekeltek abban, hogy a nyílt adatokat teljes mértékben beépíték üzleti folyamataikba. Ha az adatok rendelkezésre állnak és könnyen feldolgozhatók az ügyfél számára, akkor automatikusan ellenőrizhetők, ami kiküszöböli a mennyiségek túlbecsülésének és a becslések manipulálásának lehetőségét.

A Világgazdasági Fórum "Az építőipar jövőjének alakítása" című jelentése (2016) [5] szerint az ágazat egyik legfontosabb kihívása továbbra is az ügyfél passzív szerepe. Mindazonáltal az ügyfeleknek nagyobb felelősséget kellene vállalniuk a projektek eredményéért - a korai tervezéstől kezdve a fenntartható interakciós modellek kiválasztásán át a teljesítmény nyomon követéséig. A projektgazdák aktív részvételle nélkül az építőipar rendszerszintű átalakítása lehetetlen.

A mennyiség és a költségszámítás feletti ellenőrzés elvesztése már más iparágakat is átalakított az elmúlt 20 évben, lehetővé téve az ügyfelek számára, hogy közvetlenül, közvetítők nélkül, a oldalon keresztül akadályozzák meg céljaikat. A digitalizáció és az adatok átláthatósága számos hagyományos üzleti modellt átalakított, például a taxisoförököt az Uber megjelenésével (10.1-4. ábra), a szállodásokat az Airbnb megjelenésével, a kiskereskedőket és boltokat az Amazon felemelkedésével, a bankokat pedig a neobankok és a decentralizált fintech ökoszisztemák megjelenésével, ahol az információhoz való közvetlen hozzáférés és idő- és költségszámítások automatizálása jelentősen csökkentette a közvetítők szerepét.



Ábra: 10.1-4 Az építőiparnak ugyanazzal az überizációval kell szembenéznie, mint 10 ével ezelőtt a taxisoknak, a szállodásoknak és a kiskereskedőknek.

Az adatokhoz és a feldolgozásukhoz szükséges eszközökhöz való hozzáférés demokratizálódásának folyamata elkerülhetetlen, és idővel a nyílt adatok minden projektelemre vonatkozóan az ügyfelek igénye és új szabvány lesz. Ezért a nyílt formátumok és az átlátható számítások bevezetésének kérdéseit a befektetők, az ügyfelek, a bankok és a magántőkealapok (magántőke) fogják előmozdítani - azok, akik a megépített objektumok végfelhasználói, majd évtizedekig működtetik az objektumot.

A nagybefektetők, az ügyfelek és a bankok már most átláthatóságot követelnek az építőiparban. Az Accenture "Creating more value through capital projects" (2020) [20] című tanulmánya szerint az átlátható és megbízható adatok döntő tényezővé válnak az építőipari beruházási döntésekben. Amint arra a szakértők rámutatnak, átláthatóság nélkül lehetetlen a megbízható és hatékony projektmenedzsment, különösen válságok idején. Emellett az eszköztulajdonosok és a kivitelezők egyre inkább olyan szerződések felé mozdulnak el, amelyek ösztönzik az adatmegosztást és az együttműködésen alapuló elemzést, tükrözve a befektetők, a bankok és a szabályozó hatóságok növekvő igényeit az elszámoltathatóság és az átláthatóság iránt.

A beruházó, a megrendelő mozgása az ötlettől a kész épületig a jövőben olyan lesz, mintha robotpilótával utazna - sofőr nélkül, építőipari cég formájában, a spekulációtól és a bizonytalanságtól való függetlenséget ígéri.

A nyílt adatok és az automatizálás kora elkerülhetetlenül megváltoztatja az építőipart, ahogyan azt már megtette a banki, kereskedelmi, mezőgazdasági és logisztikai ágazatban. Ezekben az iparágakban a közvetítők szerepe és az üzletkötés hagyományos módszerei átadják helyüket az automatizálásnak és a robotizációnak, és nem hagynak teret az indokolatlan feláraknak és a spekulációról.

Az emberi gazdasági tevékenységek adatai és folyamatai nem különböznek attól, amivel az építőiparban dolgozó szakembereknek foglalkozniuk kell. Hosszú távon az építőipari vállalatok, amelyek ma az ár- és szolgáltatásminőségi szabványok meghatározásával uralják a piacot, elveszíthetik a megrendelő és az

építési projekt közötti kulcsfontosságú közvetítő szerepükét.

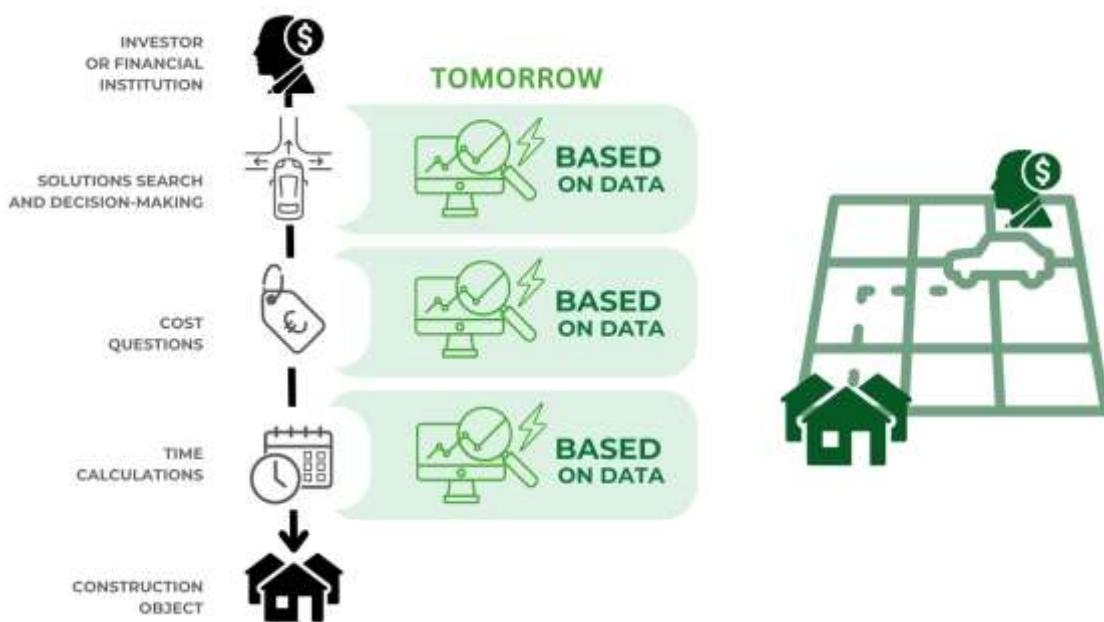
Az überizáció megoldatlan problémái mint utolsó esély az idő kihasználására az átalakulás érdekében

De tértünk vissza az építőipar realitásaihoz. Miközben a gazdaság egyes ágazataiban megjelennek az önvezető autók, a decentralizált pénzügyi rendszerek és a mesterséges intelligencián alapuló megoldások, az építőipari vállalatok jelentős része még mindig papír alapú szervezet, ahol a kulcsfontosságú döntéseket inkább az egyes szakemberek intuíciója és tapasztalata alapján hozzák meg.

Ebben a paradigmában egy modern építőipari vállalat egy 20 éves taxitársasághoz hasonlítható, amely ellenőrzi az erőforrásokat, az útvonalakat és a szállítási határidőket, és felelős az "utazás" időzítéséért és költségeiért - a projektötlettől (logisztikai és telepítési folyamat) a projekt átadásáig. Ahogyan a GPS (az építőipari IoT, az RFID) és a gépi tanulási algoritmusok az idő/költségszámításokban egykor átalakították a közlekedés területét, az adatok, az algoritmusok és az AI -ügynökök képesek átalakítani az építőipari menedzsmentet - az intuitív értékeléstől a prediktív, irányított modellekig. Az elmúlt 20 évben számos iparágban - pénzügy, mezőgazdaság, kiskereskedelem és logisztika - fokozatosan eltűnt az adatok átláthatatlanságán keresztül történő spekuláció képessége. Az árak, szállítási költségek vagy pénzügyi tranzakciók automatikusan és statisztikailag megalapozottan kerülnek kiszámításra - a digitális platformokon néhány másodperc alatt.

A jövőre nézve az építőipari vállalatoknak fel kell ismerniük, hogy az adatokhoz való hozzáférés és az azok elemzéséhez szükséges eszközök demokratizálása megzavarja a projektköltségek és -idők becslésének hagyományos megközelítését, és megszünteti a lehetőséget az átláthatatlan mennyiségi és áradatokon alapuló spekulációra.

Mint a szabályozott úton való vezetés a sofőr beavatkozása nélkül, a jövő építési folyamatai egyre inkább egy "Uberizált" rendszerhez fognak hasonlítani - automatizált idő- és költségbecsléssel, átlátható feladatirányítással és minimális emberi tényezőktől való függőséggel. Ez megváltoztatja az ötlettől a megvalósításig tartó "út" jellegét - kiszámíthatóbbá, irányíthatóbbá és adatvezérelté teszi azt.



Ábra 10.1-5 Az építés során a költségeket és az "utazási" időt a és statisztikai eszközökkel történő gépi tanulás segítségével határozzák meg.

A világ szinte minden országában fokozatosan bevezetésre kerülő új szabályozások és követelmények, amelyek kötelezik a CAD- (BIM)-modellek átadását az ügyfeleknek vagy az építési projekteket finanszírozó bankoknak, lehetőséget adnak az ügyfélnek és a megrendelőnek arra, hogy önállóan biztosítsák a költség- és terjedelem számítások átláthatóságát. Ez különösen fontos a nagy megrendelők és befektetők számára, akik megfelelő kompetenciákkal és eszközökkel rendelkeznek a volumenek gyors elemzéséhez és a piaci árak nyomon követéséhez. A nagyszabású standard projekteket - üzletek, irodaházak, lakóparkok - megvalósító vállalatoknál az ilyen gyakorlatok lassan megszokottá válnak.

Ahogy a modellek információtartalma egyre teljesebbé és szabványosabbá válik, a manipuláció és a spekuláció lehetősége gyakorlatilag megszűnik. A digitális átalakulás fokozatosan megváltoztatja az építőipar játékszabályait, és azok a vállalatok, amelyek nem alkalmazkodnak ezekhez a változásokhoz, komoly kihívásokkal nézhetnek szembe.

A fokozódó verseny, a technológiai zavarok és a csökkenő árrések hatással lehetnek az üzleti fenntarthatóságra. A korlátozott likviditás miatt egyre több iparági szereplő fordul az automatizálás, az analitika és az adatechnológiák felé, hogy javítsa a hatékonyságot és a folyamatok átláthatóságát. Ezek az eszközök fontos erőforrássá válnak ahhoz, hogy a változó gazdasági környezetben versenyképesek maradjanak.

Nem biztos, hogy érdemes megvárni, amíg a külső körülmények sürgős cselekvésre kényszerítenek - sokkal hatékonyabb, ha már ma elkezdjük a felkészülést a digitális kompetenciák megerősítésével, a modern megoldások bevezetésével és az adatközpontú kultúra kiépítésével.

Az építőipar nagyszabású digitális átalakulásának egyik utolsó kulcsfontosságú technológiai akadálya, amely az elkövetkező években minden vállalatot érinteni fog, az építési projektek elemeinek automatikus osztályozásának problémája.

Megbízható, pontos és skálázható osztályozás nélkül lehetetlen megteremteni a teljes körű analitika, a folyamatautomatizálás és az életciklus-menedzsment alapjait mesterséges intelligencia és prediktív modellek segítségével. Amíg a tárgyak osztályozása még mindig a tapasztalt szakemberek - művezetők, tervezők, becslők - kézi értelmezésén múlik, addig az építőiparnak még mindig van lehetősége. Ezt az időt ki lehet használni az elkerülhetetlen váltoásokra való felkészülésre: az átláthatóság iránti növekvő igényekre, az eszközök és adatok demokratizálására, valamint az automatikus osztályozási rendszerek megjelenésére, amelyek gyökeresen megváltoztatják a játékszabályokat.

Az építőipari világ elemeinek automatikus osztályozásának feladata összetettségét tekintve a pilóta nélküli vezetési rendszerek objektumfelismeréséhez hasonlítható, ami az egyik fő kihívás. Képzeljünk el egy pilóta nélküli autót, amely A pontból B pontba tart (10.1-5. ábra). A jelenlegi automatikus vezetési rendszerek megakadnak a lidarok és kamerák által felismert objektumok osztályozásának problémáján. Nem elég, ha az autó egyszerűen "lát" egy akadályt vagy tereptárgyat, hanem fel kell tudnia ismerni, hogy az egy gyalogos, egy közlekedési tábla vagy egy szemetes.

Hasonló alapvető kihívás előtt áll az építőipar egésze. A projektelemek - például ablakok, ajtók vagy oszlopok - dokumentációban rögzíthetők, CAD-modellekben ábrázolhatók, az építkezésen lefényképezhetők vagy lézerszkennelésből származó pontfelhőkben felismerhetők. Vizuális vagy durva geometriai felismerésük azonban nem elegendő egy valóban automatizált projektmenedzsment rendszer kiépítéséhez. Biztosítani kell, hogy minden egyes elemet pontosan és következetesen kategorizáljanak egy olyan típusba, amely minden későbbi folyamat során - a becslésekktől és a specifikációktól kezdve a logisztikán és a készletellenőrzésen át a legfontosabb, az üzemeltetésig - egyértelműen azonosítható (4.2-6. ábra).

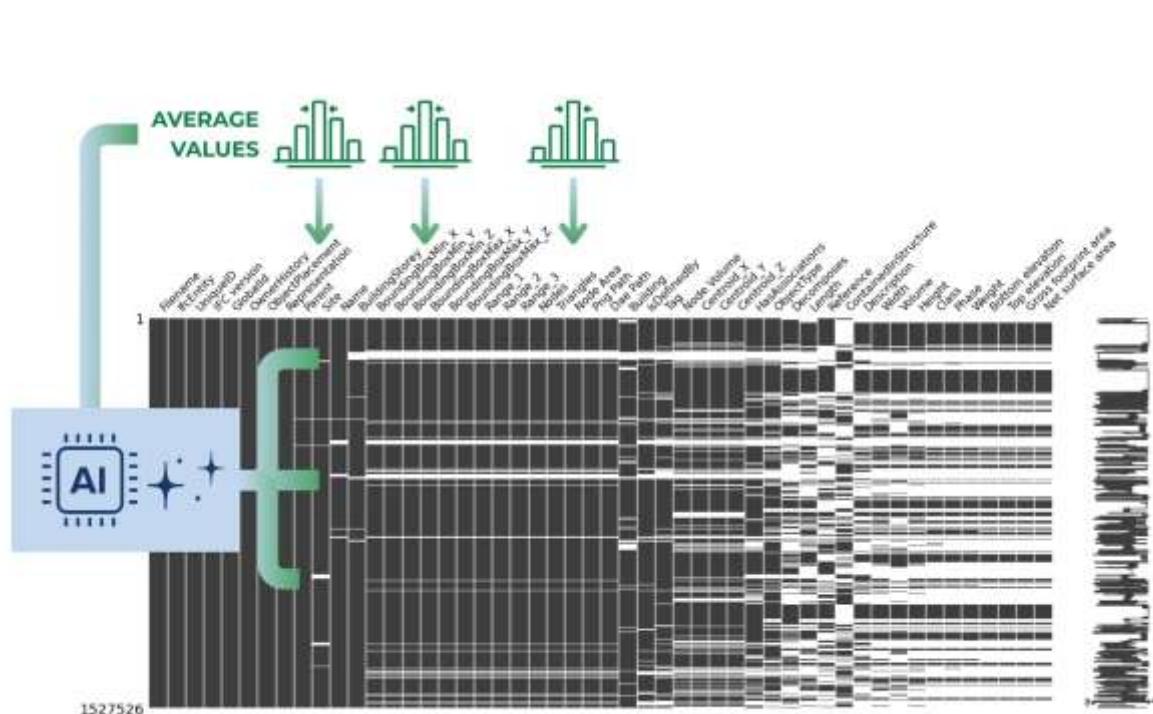
Ebben a szakaszban - a felismerésről az értelmes osztályozásra való áttérésnél - merül fel az egyik legfontosabb akadály. Még ha a digitális rendszerek technikailag képesek is megkülönböztetni és azonosítani a modellekben és az építkezésen található objektumokat, a fő nehézség az elemtípus helyes és kontextuálisan stabil meghatározása a különböző szoftverkörnyezetekben.. Például egy ajtót a tervező a CAD-modellben "ajtó" kategóriájú elemként jelölhet meg, de az ERP- vagy PMIS-rendszerbe való átvitelkor a tervező hibájából vagy a rendszerek közötti következetlenségek miatt tévesen tipizálódhat. Ezenkívül az elem gyakran elveszít néhány fontos attribútumot, vagy eltűnik a rendszer könnyeliséből az adatexport és -import során. Ez az adatáramlásban hiányosságot eredményez, és aláássa az építési folyamatok végponttól végpontig tartó digitalizálásának elvét. Ez kritikus szakadékot teremt a "látható" és az "érthető" szemantikai jelentés között, ami aláássa az adatok integritását, és jelentősen megnehezíti a folyamatok automatizálását az építési projekt teljes életciklusa során.

A építőelemek univerzális osztályozásának megoldása a big data és a gépi tanulási technológiák segítségével (10.1-6. ábra) az egész iparát érintő átalakulás katalizátora lesz - és talán sok építőipari vállalat számára váratlan felfedezés. Egy egységes, tanulható osztályozási rendszer lesz az alapja a skálázható analitikának, a digitális irányításnak és a mesterséges intelligencia bevezetésének az építőipari szervezetek minden napjai gyakorlatába.

Az NVIDIA és más technológiai vezetők már most is kínálnak olyan megoldásokat más iparágakban, amelyek képesek automatikusan kategorizálni és strukturálni hatalmas mennyiségű szöveges és vizuális információt.

Az NVIDIA NeMo Curator modellje [161] például az adatok automatikus osztályozására és előre meghatározott kategóriákba történő kategorizálására specializálódott, és kulcsszerepet játszik az információfeldolgozó pipelinek optimalizálásában a generatív AI modellek finomhangolásához és előzetes betanításához. A Cosmos platformot valós videókon és 3D-s jeleneteken képzik [162], ami alapot nyújt az NVIDIA ökoszisztémájában már épülő autonóm rendszerek és digitális ikerk számára. Az NVIDIA Omniverse, amely 2025-re az USD formátummal való munkavégzés vezető eszközévé vált, amely egy univerzális jelenetleírás, amely idővel felválthatja az IFC formátumot a tervezési információátviteli folyamatokban. Az Isaac Sim robotikai folyamatszimulátorral [163] együtt az olyan megoldások, mint a NeMo Curator, a Cosmos és az Omniverse az automatizálás új szintjét képviselik: az adattisztítástól és -szűréstől a gyakorlóhalmozok generálásáig, az objektumtulajdonságok modellezéséig és az építkezésen történő robotkiképzésig. Ráadásul ezek az eszközök mindegyike ingyenes és nyílt forráskódú, ami jelentősen csökkenti az elfogadás akadályait a mérnöki és építőipari gyakorlatban.

Az adatok automatikus osztályozása strukturált táblázatok szintjén nem olyan nehéz, mint amilyennek első pillantásra tűnik. Amint azt az előző fejezetben bemutattuk (9.1-10. ábra), a hiányzó vagy hibás osztályértékek pótlása más elemek hasonló paraméterei alapján lehetséges, ha vannak felhalmozott múltbeli adatok. Ha hasonló jellemzőkkel rendelkező elemeket már több befejezett projektben is helyesen osztályoztak, a rendszer nagy valószínűséggel tud megfelelő értéket javasolni egy új vagy hiányos elemhez (10.1-6. ábra). Az ilyen, átlagolt értékeken és kontextuselemzésen alapuló logika különösen hatékony lehet a becslésekkel, specifikációkból vagy CAD-modellekkel származó táblázatos adatok tömeges feldolgozásakor.

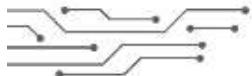


Ábra 10.1-6 A gépi tanulás segíthet automatikusan átlagos értékeket találni a kitöltetlen (fehér dobozok) táblázat paramétereihez a múltbeli projektek alapján.

A gépi tanulás ilyen gyors fejlődésének fényében egyértelmű: 2025-ben naivitás lenne azt hinni, hogy az épületelemek automatikus osztályozásának problémája sokáig megoldatlan marad. Igen, a jelenlegi algoritmusok még nem érték el a teljes érettséget, különösen a hiányos vagy heterogén adatok esetében, de az alkalmazkodási lehetőség gyorsan bezárul.

Előnyben lesznek azok a vállalatok, amelyek már befektetnek adataik összegyűjtésébe, tisztításába és rendszerezésébe, valamint az ETL automatizálási eszközök bevezetésébe. A többiek azt kockáztatják, hogy lemaradnak - ahogyan a vállalatok egykor a közlekedési és a pénzügyi ágazatban sem tudtak megfelelni a digitális átalakulás kihívásainak.

Azok, akik továbbra is a kézi adatkezelésre és a költség- és időbecslés hagyományos módszereire hagyatkoznak, azt kockáztatják, hogy a 2000-es évek taxiflottáinak helyzetébe kerülnek, és a 2020-as évek elejére nem tudnak alkalmazkodni a mobilalkalmazások és az automatizált útvonalszámítások korszakához.



FEJEZET 10.2.

GYAKORLATI ÚTMUTATÓ AZ ADATVEZÉRELT MEGKÖZELÍTÉS MEGVALÓSÍTÁSHOZ

Az elmélettől a gyakorlatig: Útiterv az építőipar digitális átalakulásához

Az építőipar fokozatosan új fejlődési szakaszba lép, ahol a megszokott folyamatokat egyre inkább digitális platformok és átlátható interakciós modellek egészítik ki, sőt néha fel is váltják. Ez nemcsak kihívások elő állítja a vállalatokat, hanem jelentős lehetőségek elő is. Azok a szervezetek, amelyek már ma hosszú távú digitális stratégiát építenek ki, nem csak megőrizhetik piaci pozíciójukat, hanem ki is bővíthetik azt az által, hogy korszerű megközelítéseket és megbízható, technológiailag támogatott megoldásokat kínálnak az ügyfeleknek.

Fontos felismerni, hogy a fogalmak és technológiák ismerete csak kiindulópont. A vezetők és szakemberek gyakorlati kérdéssel szembesülnek: hol kezdjék el a megvalósítást, és hogyan alakítsák át az elméleti ötleteket valódi értékké. Emellett egyre gyakrabban merül fel a kérdés: mire épül majd az üzlet, ha a hagyományos költségszámítási és időzítési módszereket a megrendelő bármikor felülvizsgálhatja.

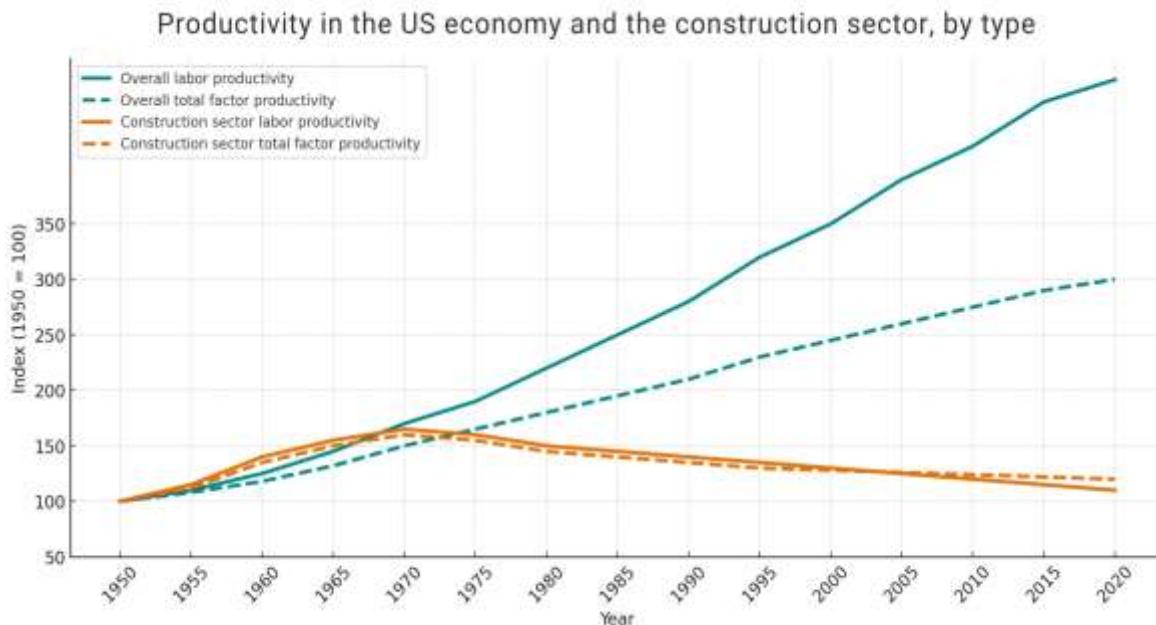
A válasz valószínűleg nem annyira a technológiában rejlik, hanem egy új szakmai kultúra kialakításában, ahol az adatokkal való munkát a minden napjai gyakorlat szerves részének tekintik. A digitális technológia és az innováció iránti figyelem hiánya oltotta be az építőipart abba a súlyos lemaradásba, amely az elmúlt évtizedekben megfigyelhető volt [43].

A McKinsey szerint az építőiparban a K+F-kiadások a bevétel kevesebb mint 1%-át teszik ki, míg az autóiparban és a repülőgépiparban ez az arány eléri a 3,5-4,5%-ot. Hasonlóképpen, az építőiparban az IT-költségek az összes bevétel kevesebb mint 1%-át teszik ki [107].

Ennek eredményeként nemcsak az automatizálás szintje, hanem a munkatermelékenység is csökken az építőiparban, és 2020-ra az építőipari munkás kevesebbet termel, mint fél évszázaddal ezelőtt (10.2-1. ábra)

Az építőipari ágazat ilyen jellegű termelékenységi problémái a legtöbb fejlett és fejlődő országban egyaránt tapasztalhatók (az OECD 29 országából 16-ban csökkent az építőipar termelékenysége (2.2-1. ábra)), és nemcsak a technológia hiányára utalnak, hanem arra is, hogy rendszerszintű változásokra van szükség a vezetés, a képzés és az innováció megközelítésében.

A digitális átalakulás sikere nem annyira az eszközök számától és elérhetőségtől függ, hanem attól, hogy a szervezetek képesek-e felülvizsgálni folyamataikat és kialakítani egy olyan kultúrát, amely nyitott a változásra. Nem maga a technológia a kulcs, hanem azok az emberek és folyamatok, amelyek biztosítják annak hatékony használatát, támogatják a folyamatos tanulást és ösztönzik az új ötletek elfogadását.



10.2-1 A munkatermelékenység és a teljes erőforrás-termelékenység paradoxona az USA gazdaságában és az építőiparban (1950-2020) ([43] alapján).

A könyv korai részeiben az üzleti környezet modelljét egy erdei ökoszisztemához hasonlították (2.1-2. ábra, 1.2-4. ábra, 1.3-2. ábra). Egy egészséges erdőben az időszakos tüzek - minden pusztító erejük ellenére - kulcsszerepet játszanak a hosszú távú megújulásban. Megszabadítják a talajt a régi növényzettől, visszaadják a tárolt tápanyagokat, és helyet teremtenek az új élet számára. Egyes növényfajok még úgy is fejlődtek, hogy magjaik csak akkor nyílnak ki, ha magas tűzhőmérsékletnek vannak kitéve - ez egy természetes mechanizmus, amely ideális időt biztosít a csírázáshoz.

Hasonlóképpen az üzleti életben a válságok "ellenőrzött kiégésként" működhetnek, és olyan új megközelítések és vállalatok kialakulását ösztönzik, amelyek nem kötődnek elavult rendszerekhez. Az ilyen időszakok a nem hatékony gyakorlatok elhagyására kényszerítenek, és felszabadítják az erőforrásokat az innováció számára. Ahogyan egy tűzvész utáni erdő is úttörő növényekkel indul, úgy egy válság után egy vállalkozás is új, rugalmas folyamatokat alakít ki, amelyek egy érett információs környezet alapjává válnak.

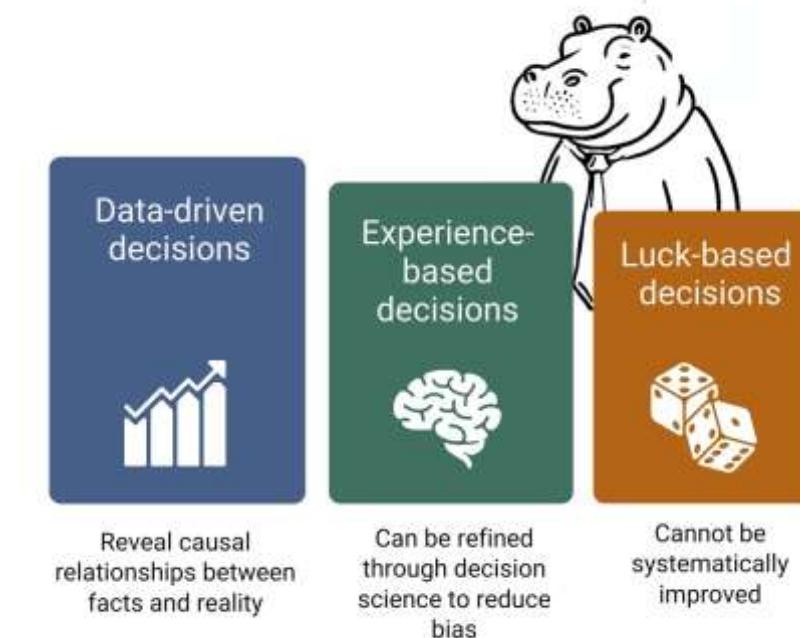
Azok a vállalatok, amelyeknek sikerül helyesen értelmezniük ezeket a "jelzőtüzekeket", és a bomlasztó energiát konstruktív változássá alakítaniuk, a teljesítmény új szintjét érik el - átláthatóbb, alkalmazkodóbb adatfolyamatokkal, amelyek fokozzák a szervezet természetes megújulási és növekedési képességét.

A mesterséges intelligencia és a gépi tanulás növekvő befolyása az üzleti környezetre már nem kétséges. Ez nem csupán átmeneti trend, hanem stratégiai szükségszerűség. Azok a vállalatok, amelyek figyelmen kívül hagyják a mesterséges intelligenciát, a versenyképesség elvesztését kockáztatják egy olyan piacon, amely egyre inkább ösztönzi az innovációt és a rugalmasságot.

A jövő azoké, akik a mesterséges intelligenciát nem csupán eszköznek tekintik, hanem lehetőségeknek arra, hogy újragondolják vállalkozásuk minden aspektusát - a folyamatok optimalizálásától a vezetői döntések meghozataláig.

A digitális alapok megteremtése: 1-5 lépés a digitális érettség felé

Ebben a fejezetben megvizsgáljuk a digitális átalakulás ütemtervét, és meghatározzuk az adatvezérelt megközelítés megvalósításához szükséges legfontosabb lépéseket, amelyek segíthetnek mind a vállalati kultúra, mind a vállalat információs ökoszisztémájának átalakításában.



10.2-2 ábra Ellenőrzött frissítés és stratégiaválasztás: eset, tapasztalat vagy adat.

A McKinsey "Why Digital Strategies Fail" (2018) című tanulmánya szerint legalább öt oka [164] van annak, hogy a vállalatok miért nem érik el a digitális átalakulás céljait.

- **Homályos definíciók:** A vezetők és a menedzserek eltérően értelmezik a "digitális" fogalmát, ami félreértekhez és ellentmondásokhoz vezet.
- **A digitális gazdaság félreértése:** sok vállalat alábecsüli a digitalizáció által az üzleti modellekben és az iparág dinamikájában okozott változások mértékét (10.1-6. ábra).
- **Az ökoszisztémák figyelmen kívül hagyása:** a vállalatok az egyedi technológiai megoldásokra (adatsílokra) összpontosítanak, figyelmen kívül hagyva a szélesebb digitális ökoszisztémákba való integráció szükségességét (2.2-2-2. ábra, 4.1-12. ábra).
- **A versenytársak digitalizációjának alábecsülése:** a vezetők nem veszik figyelembe, hogy a versenytársak is aktívan alkalmazzák a digitális technológiákat, ami a versenyelőny elvesztéséhez vezethet.
- **A digitalizáció kettősségeinek figyelmen kívül hagyása:** a vezérigazgatók a digitális átalakítás felelősségeit más vezetőkre ruházzák át, ami bürokratizálja az ellenőrzést és lassítja a változás folyamatát.

E kihívások kezelése a digitális stratégiák világos megértését és összehangolását igényli a szervezet minden szintjén. A digitális stratégia kialakítása előtt fontos megérteni a kiindulópontot. Sok szervezet

hajlamos új eszközöket és platformokat bevezetni anélkül, hogy teljes képet kapna a jelenlegi állapotról.

1. lépés: Végezze el a jelenlegi rendszerek és adatok auditálását.

A folyamatok megváltoztatása előtt fontos megérteni, hogy mi van már a helyén. Az audit elvégzése lehetővé teszi az adatkezelés gyenge pontjainak azonosítását, és annak megértését, hogy milyen erőforrások használhatók fel. Az audit egyfajta "röntgenfelvétel" az üzleti folyamatokról. Lehetővé teszi a kockázati területek azonosítását, és annak meghatározását, hogy mely adatok kritikusak a projekt vagy az üzleti tevékenység szempontjából, és melyek másodlagosak.

Fő tevékenységek:

- Térképezze fel az informatikai környezetét (Draw.io, Lucidchart, Miro, Visio vagy Canva segítségével). Sorolja fel a folyamataiban részt vevő, használt rendszereket (ERP, CAD, CAFM, CPM, SCM és egyéb), amelyeket a "Technológiák és irányítási rendszerek a modern építőiparban" című fejezetben tárgyaltunk (1.2-4. ábra)
- Értékelje az egyes rendszerek adatminőségi problémáit a duplikációk, az esetlegesen hiányzó értékek és a formátumbeli következetlenségek gyakorisága szempontjából.
- Azonosítsa a "fájdalmas pontokat" - azokat a helyeket, ahol a folyamatok megszakadhatnak, vagy gyakran manuális beavatkozást igényelnek - import, export és további érvényesítési folyamatok.

Ha azt szeretné, hogy a csapat megbízzon a jelentésekben, már a kezdetektől fogva meg kell győződni arról, hogy az adatok helyesek.

A minőségi adataudit megmutatja, hogy milyen adatok:

- További fejlesztésre szorul (automatikus tisztítási folyamatokat vagy további átalakításokat kell létrehozni).
- Ezek "szemét", amelyek csak eltömítik a rendszereket, és megszabadulhatunk tőlük, ha nem használjuk őket többé a folyamatokban.

Lehetőség van arra, hogy egy ilyen auditot saját maga végezzen el. Néha azonban hasznos egy külső tanácsadó bevonása - különösen más iparágakból: a friss szemlélet és az építőipari "sajátosságoktól" való függetlenség segít józanul értékelni a jelenlegi helyzetet, és elkerülni a bizonyos megoldások és technológiák iránti elfogultság tipikus buktatóit.

2. lépés: Az adatharmonizáció kulcsfontosságú szabványainak meghatározása.

Az auditálás után közös szabályokat kell alkotni az adatokkal való munkavégzésre. Amint azt a "Szabványok: A véletlenszerű fájloktól az intelligens adatmodellig" című fejezetben tárgyaltuk, ez segít megszüntetni a silózott adatáramlásokat.

Egyeséges szabvány nélkül minden csapat továbbra is a "saját módján" fog dolgozni, és az integrációk "állatkertjét" fogja fenntartani, ahol az adatok minden egyes konverzióval elvesznek.

Fő tevékenységek:

- Válassza ki a rendszerek közötti információcseréhez szükséges adatszabványokat:
 - A táblázatos adatok esetében ez lehet strukturált formátum, mint például a CSV, XLSX vagy hatékonyabb formátumok, mint például a Parquet.
 - Lazán strukturált adatok és dokumentumok cseréjéhez: JSON vagy XML
- Az adatmodellekkel való munka elsajátítása:
 - Kezdje a feladatok paraméterezésével a fogalmi adatmodell szintjén - az "Adatmodellezés: fogalmi, logikai és fizikai modell" című fejezetben leírtak szerint (4.3-2. ábra).
 - Ahogy egyre mélyebbre merül az üzleti folyamat logikájában, lépjön át a követelmények formalizálására a logikai és fizikai modellek paramétereinek segítségével (4.3-6. ábra).
 - A kulcsfontosságú entitások, attribútumaik és a folyamatokon belüli kapcsolataik azonosítása, és ezen kapcsolatok vizualizálása - minden az entitások, minden a paraméterek között (4.3-7. ábra).
- Használjon reguláris kifejezéseket (RegEx) az adatok érvényesítésére és szabványosítására (4.4-7. ábra), amint azt a "Strukturált követelmények és a RegEx reguláris kifejezések" című fejezetben tárgyalunk. A RegEx nem összetett, de rendkívül fontos téma a fizikai adatmodellek szintjén történő követelményalkotás munkájában.

Adatszintű szabványok és a folyamatok vizualizálása nélkül lehetetlen következetes és skálázható digitális környezetet biztosítani. Ne feledje, hogy "a rossz adat drága". A hiba költsége pedig annál nagyobb, minél összetettebbé válik egy projekt vagy szervezet. A formátumok egységesítése, az elnevezési, szerkezeti és érvényesítési szabályok meghatározása befektetés a jövőbeli megoldások stabilitásába és skálázhatóságába.

3. lépés: DataOps bevezetése és a folyamatok automatizálása.

Jól meghatározott architektúra nélkül a vállalatoknak elkerülhetetlenül szembe kell nézniük a silózott információs rendszerekben tárolt, eltérő adatokkal. Az adatok nem lesznek integrálva, több helyen duplikálódnak, és karbantartásuk költséges lesz.

Képzeljük el, hogy az adatok víz, az adatarchitektúra pedig a csővezetékek összetett rendszere, amely a vizet a tárolási forrástól a felhasználási pontig szállítja. Az adatarchitektúra határozza meg, hogy az információkat hogyan gyűjtik, tárolják, alakítják át, elemzik és juttatják el a végfelhasználókhoz vagy alkalmazásokhoz.

A DataOps (Data Operations) egy olyan módszertan, amely az adatok gyűjtését, tisztítását, érvényesítését és felhasználását egyetlen automatizált folyamatfolyamatba integrálja, amint azt a könyv 8. részében részletesen tárgyalunk.

Fő tevékenységek:

- Létrehozza és testre szabja az ETL -szállítót a folyamatok automatizálása érdekében:

- Extract: automatikus adatgyűjtés megszervezése PDF dokumentumokból (4.1-2 ábra, 4.1-5 ábra, 4.1-7 ábra), Excel táblázatokból, CAD -modellek ből (7.2-4 ábra), ERP - rendszerekből és más forrásokból, amelyekkel Ön dolgozik.
- Transform: automatikus folyamatok beállítása az adatok egységes strukturált formátumba történő átalakítására és a zárt alkalmazásokon kívül zajló számítások automatizálására (7.2-8. ábra).
- Betöltés: próbáljon meg automatikus adatfeltöltést létrehozni összefoglaló táblázatokba, dokumentumokba vagy központi tárolókba (7.2-9 ábra, 7.2-13 ábra, 7.2-16 ábra).
- Automatizálja a számítási és QTO (Quantity Take-Off) folyamatokat, ahogyan azt a "QTO Quantity Take-Off: Projektadatok csoportosítása attribútumok szerint" című fejezetben tárgyaltuk:
 - Állítsa be a térfogatok automatikus extrahálását a CAD -modellek ből, API-k, beépülő modulok vagy reverse engineering eszközök segítségével (5.2-5. ábra).
 - Hozzon létre szabályokat a különböző osztályok elemeinek attribútumok szerinti csoportosítására táblázatok formájában (5.2-12 ábra).
 - Próbálja meg automatizálni a moduláris zárt rendszereken kívüli, gyakran ismétlődő mennyiség- és költségszámításokat (5.2-15. ábra).
- Kezdje el használni a Python és a Pandas oldalakat az adatok feldolgozásához, ahogyan azt a "Python Pandas: nélkülvilágos eszköz az adatokkal való munkához" című fejezetben tárgyaltuk:
 - DataFrame alkalmazása az XLSX fájlokkal való munkához és a táblázatos adatok feldolgozásának automatizálásához (3.4-6. ábra).
 - Az információk összesítésének és átalakításának automatizálása különböző Python könyvtárak segítségével
 - Az LLM használatával egyszerűsítheti a kész kódblokkok és teljes Pipelines írását (7.2-18 ábra).
 - Próbáljon ki egy csővezetéket építeni Python nyelven, amely hibákat talál vagy rendellenességeket észlel, és értesítést küld a felelős személynek (pl. a projektvezetőnek) (7.4-2. ábra).

A DataOps-elveken alapuló automatizálás lehetővé teszi, hogy a kézi és töredékes adatkezelésről fenntartható és megismételhető folyamatokra térjen át. Ez nem csak csökkenti az alkalmazottak terheit, akik nap mint nap ugyanazokkal az átalakításokkal foglalkoznak, hanem drámaian növeli a teljes információs rendszer megbízhatóságát, skálázhatóságát és átláthatóságát is.

4. lépés: Nyílt adatkezelési ökoszisztemája létrehozása.

A zárt moduláris rendszerek fejlődése és az új eszközökkel való integrációjuk ellenére a vállalatok komoly problémával szembesülnek - az ilyen rendszerek növekvő komplexitása meghaladja hasznosságukat. A kezdeti elképzelés, miszerint egyetlen, minden üzleti folyamatot lefedő, szabadalmaztatott platformot kell létrehozni, túlzott központosításhoz vezetett, ahol minden változtatás jelentős erőforrásokat és időt igényel az alkalmazkodáshoz.

Amint azt a "Vállalati micélium: Hogyan kapcsolják össze az adatok az üzleti folyamatokat" című fejezetben tárgyaltuk, a hatékony adatkezeléshez olyan nyitott és egységes ökoszisztémára van szükség, amely összekapcsolja az összes információforrást.

Az ökoszisztéma kulcsfontosságú elemei:

■ Válassza ki a megfelelő adattárolót:

- A táblázatokhoz és számításokhoz használjon adatbázisokat - például PostgreSQL vagy MySQL (3.1-7. ábra).
- Dokumentumok és jelentések esetében a felhőalapú tárolás (Google Drive, OneDrive) vagy a JSON formátumot támogató rendszerek alkalmasak lehetnek.
- Tekintse meg az adattárházak, adattavak és más, nagy mennyiséggű információ központi tárolására és elemzésére szolgáló eszközök képességeit (8.1-8. ábra).

■ Megoldások bevezetése a védett adatokhoz való hozzáférésre:

- Ha saját rendszereket használ, konfigurálja az API-n vagy SDK-n keresztül történő hozzáférést a külső feldolgozásra szánt adatok megszerzéséhez (4.1-2. ábra).
- Ismerkedjen meg a reverse engineering eszközök lehetőségeivel a CAD formátumokhoz (4.1-13. ábra).
- ETL-csatornák beállítása, amelyek időszakosan adatokat gyűjtenek az alkalmazásokból vagy szerverekből, átalakítják azokat nyílt, strukturált formátumba, és tárolókba mentik (7.2-3. ábra).
- A csapaton belüli megbeszélés arról, hogy hogyan lehet az adatokhoz való hozzáférést saját szoftver nélkül biztosítani.
- Ne feledje: az adatok fontosabbak, mint az interfészek. Az információk struktúrája és elérhetősége, nem pedig a konkrét felhasználói interfész eszközök jelentik a hosszú távú értéket.

■ Gondolkodjon el egy kiválósági központ (CoE) létrehozásán az adatokkal kapcsolatban, ahogyan azt az "Adatmodellezés kiválósági központja (CoE)" fejezetben tárgyaltuk, vagy hogyan tudna más módon adatszakértelmet biztosítani (4.3-9. ábra).

Az adatkezelési ökoszisztéma olyan egységes információs teret hoz létre, amelyben a projekt valamennyi résztvevője egységes, naprakész és ellenőrzött információkkal dolgozik. Ez az alapja a skálázható, rugalmas és megbízható digitális folyamatoknak

Az adatokban rejlő lehetőségek felszabadítása: 5-10 lépés a digitális érettség felé

A technikai integráció mellett a digitális megoldások sikeres bevezetésének fontos tényezője a végfelhasználók általi elfogadásuk. Az ügyfelek vagy felhasználók bevonása a teljesítménymérésbe egyszerre jelenti a felhasználói élmény javításának és a vállalaton belüli változások kezelésének kihívását. Ha egy megoldás nem illeszkedik a megszokott munkafolyamatba, vagy nem old meg valós felhasználói vagy ügyfélproblémákat, akkor nem fogják használni, és ezt semmilyen kiegészítő intézkedés és ösztönző

nem fogja helyrehozni.

Az átalakítás egy iteratív folyamat, amely a felhasználói interakciós adatok elemzésén alapul, gyakori tesztelési ciklusokkal, folyamatos visszajelzésekkel és finomításokkal.

5. lépés: Adatkultúra kialakítása, a személyzet képzése és visszajelzések gyűjtése

Még a legfejlettebb rendszer sem működik a munkavállalók elkötelezettsége nélkül. Olyan környezetet kell teremtenie, ahol az adatokat napi szinten használják, és a csapat megérti az értékét.

Az Egyesült Királyság kormánya által közzétett 2024. évi "Data Analytics and AI in Government Project Delivery" című jelentés [83] megjegyzi, hogy az adatfeldolgozáshoz és -értelmezéshez szükséges kompetenciákkal rendelkező szakemberek képzése kritikus fontosságú a adatelemzés és az AI sikeres megvalósítása szempontjából.

Az adatelemzési szakértelem hiánya az egyik legfontosabb probléma, amely korlátozza a digitális átalakulást. A vezetők hozzászoktak a bevett rutinokhoz: negyedéves ciklusokhoz, rangsorolt kezdeményezésekhez és a projektek előrevitelének hagyományos módjaihoz. A változáshoz markáns vezetőre van szükség - aki elég magas rangú ahhoz, hogy befolyással rendelkezzen, de nem olyan magas, hogy legyen ideje és motivációja egy hosszú távú átalakítási projekt vezetésére.

Fő tevékenységek:

- Annak felismerése, hogy a magasan fizetett alkalmazottak véleményén alapuló szubjektív döntésekéről (HiPPO) át kell térni a tényeken és adatokon alapuló döntéshozatali kultúrára, ahogyan azt a "HiPPO vagy a vélemény veszélye a döntéshozatalban" című fejezet tárgyalja (2.1-9. ábra).
- Rendszeres képzés szervezése:
 - Tartsanak képzéseket a strukturált adatok használatáról, és hívjanak meg olyan szakértőket más iparágakból, akik nem elfogultak a ma az építőiparban népszerű termékek és koncepciók iránt.
 - Adatelemzési megközelítések és eszközök megvitatása kollégáival, valamint önállóan tanulja meg a gyakorlati munkát olyan eszközökkel, mint a Python, a pandas és az LLM (4.1-3. ábra, 4.1-6. ábra).
 - Hozzon létre egy oktatónövekvőt (lehetőleg rövid videókkal) az adatok strukturálása (3.2-15. ábra) és az adatmodellek létrehozása (4.3-6. ábra, 4.3-7. ábra) téma körében.
- A modern tanulási technológiák felhasználása:
 - Nyelvi modellek (LLM) használata a kód- és adatmanipuláció támogatására, beleértve a kódgenerálást, a refaktorálást és az elemzést, valamint a táblázatos információk feldolgozását és értelmezését (3.4-1 ábra).
 - Fedezze fel, hogy az LLM által generált kód hogyan adaptálható és integrálható egy teljes csővezeték-megoldásba, amikor offline fejlesztőkörnyezetben (IDE) dolgozik (4.4-14. ábra, 5.2-13. ábra).

Ha egy vezető továbbra is a "régimódi módon" hoz döntéseket, semmilyen képzés nem fogja meggyőzni az embereket arról, hogy komolyan vegyék az analitikát.

Az adatkultúra kiépítése lehetetlen folyamatos visszajelzés nélkül. A visszajelzés lehetővé teszi a folyamatok, eszközök és stratégiák hiányosságainak azonosítását, amelyek nem fedezhetők fel belső jelentések vagy hivatalos KPI-k mérőszámok segítségével. Az Ön megoldásainak felhasználóitól származó kiegészítő megjegyzések nem nyújtanak gyakorlati értéket. A kritikus visszajelzés az értékes, különösen, ha az konkrét megfigyeléseken és tényeken alapul. Az ilyen információk megszerzése azonban erőfeszítést igényel: olyan folyamatokat kell kiépítenie, amelyekben a résztvevők - belső és külső - torzítás nélkül oszthatják meg észrevételeiket (lehet, hogy érdemes névtelenül tenni), és nem kell attól tartania, hogy véleményük hatással lehet saját munkájukra. Fontos, hogy ezt torzítás nélkül és a saját magukra nézve negatív következményektől való félelem nélkül tegyék.

Minden tanulás végső soron önképzés [165].

- Milton Friedman, amerikai közgazdász és statisztikus

Az analitikai eszközök bevezetését a gyakorlati hatékonyságuk rendszeres ellenőrzésének kell kísérnie (ROI, KPI-k), ami csak az alkalmazottak, az ügyfelek és a partnerek strukturált visszajelzésével érhető el. Ez nemcsak a hibák megismétlődésének elkerülését teszi lehetővé, hanem azt is, hogy a vállalatok gyorsabban alkalmazkodjanak a környezet változásaihoz. A visszajelzések gyűjtésére és elemzésére szolgáló mechanizmus megléte a szervezet érettségének egyik jele, amely az alkalmi digitális kezdeményezésekkel a folyamatos fejlesztés fenntartható modellje felé halad (2.2-5. ábra).

6. Lépés: A kísérleti projektektől a méretnövelésig

Válassza ki a csatákat, amelyek elég nagyok ahhoz, hogy számítanak, és elég kicsik ahhoz, hogy megnyerje őket.

- Jonathan Kozol

A digitális átalakulás "egyszerre és mindenhol" történő elindítása rendkívül kockázatos. Hatékonyabb megközelítés a kísérleti projektekkel való kezdés és a sikeres gyakorlatok fokozatos kiterjesztése.

Fő tevékenységek:

- Válassza ki a megfelelő projektet a kísérleti projekthez:
 - Határozzon meg egy konkrét üzleti célt vagy folyamatot mérhető eredményekkel (KPI, ROI) (7.1-5. ábra).

- Válasszon ki egy ETL-automatizálási folyamatot, például automatikus adatérényesítést vagy munkaterhelés-számítást (QTO) a Python és Pandas segítségével (5.2-10. ábra)
- Egyértelmű mérőszámok meghatározása a sikerhez (pl. a vizsgálati specifikációk vagy az adatérényesítési jelentések elkészítésének idejét egy hétről egy napra csökkenteni).

■ Alkalmazzon iteratív megközelítéseket:

- Kezdje egyszerű adatkonverziós folyamatokkal, és hozza létre a több formátumú adatok áramló konverzióját a folyamataihoz szükséges formátumokra (4.1-2. ábra, 4.1-5. ábra).
- Fokozatosan növelte a feladatok összetettségét és bővíse a folyamatok automatizálását az IDE-ben dokumentált kódblokkokon alapuló teljes Pipeline kialakításával (4.1-7. ábra, 7.2-18. ábra).
- Dokumentálja és rögzítse (lehetőleg rövid videókkal) a sikeres megoldásokat, és ossza meg azokat a kollégákkal vagy a szakmai közösségekben.

■ Sablonok és kísérő dokumentáció kidolgozása az ilyen megoldások megismétlésére, hogy kollégái (vagy a szakmai közösség tagjai, beleértve a közösségi médiafelhasználókat is) hatékonyan használhassák azokat.

A lépésről lépésre történő "roll-up" lehetővé teszi, hogy megőrizze a változások magas minőségét, és ne essen a párhuzamos megvalósítások káoszába. A "kicsiktől a nagyokig" stratégia minimalizálja a kockázatokat, és lehetővé teszi, hogy tanuljon a kis hibákból anélkül, hogy azok kritikus problémává nőnének.

A projektalapú megközelítésről - amelyben a munkavállalók csak részben vesznek részt - az állandó csapatok (pl. szakértői központok) létrehozására való áttérés segít biztosítani a termék fenntartható fejlesztését az első verzió kiadása után is. Az ilyen csapatok nemcsak támogatják a meglévő megoldásokat, hanem továbbfejlesztik azokat.

Ez csökkenti a hosszadalmas jóváhagyásoktól való függőséget: a csapattagok felhatalmazást kapnak arra, hogy saját felelősségi körükön belül hozzanak döntéseket. Ennek eredményeképpen a vezetők mentesülnek a mikromenedzsment szükségessége alól, a csapatok pedig a valódi értékteremtésre összpontosítanak.

Az új megoldások kifejlesztése nem sprint, hanem maraton. Azok járnak sikerrel, akik kezdetben hosszú távú, következetes munkára törekednek.

Fontos felismerni, hogy a technológia folyamatos fejlesztést igényel. A technológiai megoldások hosszú távú fejlesztésébe való befektetés a sikeres működés alapja.

7. lépés: Nyílt adatformátumok és megoldások használata

Amint azt a moduláris platformokról (ERP, PMIS, CAFM, CDE stb.) szóló fejezetekben tárgyaltuk, fontos a nyílt és univerzális adatformátumokra összpontosítani, amelyek biztosítják a szállítói megoldásoktól való függetlenséget és növelik az információk hozzáférhetőségét minden érdekelt fél számára.

Fő tevékenységek:

■ A zárt formátumokról a nyílt formátumok felé való elmozdulás:

- Használjon nyílt formátumokat a védett formátumok helyett, vagy találjon módot a zárt formátumok automatikus feltöltésére vagy nyílt formátummá alakítására (3.2-15. ábra).
- A Parquet, CSV, JSON, XLSX, amelyek a legtöbb modern rendszer közötti csere szabványok (8.1-2. ábra).
- Ha a 3D geometriával való munka fontos szerepet játszik a folyamataiban, fontolja meg az olyan nyílt formátumok használatát, mint az USD, glTF, DAE vagy OBJ (3.1-14 ábra).

■ Vektoros adatbázisok használata adatbázisok hatékony elemzéséhez és információkereséshez:

- Használja a Bounding Box és más módszereket a 3D -geometria egyszerűsítésére (8.2-1 ábra).
- Gondolja át, hogy hol tudja megvalósítani az adatvektorizálást - szövegek, objektumok vagy dokumentumok numerikus ábrázolássá alakítását (8.2-2. ábra).

■ Alkalmazza a nagy adatalemzési eszközöket:

- A felhalmozott múltbeli adatok (pl. PDF, XLSX, CAD) tárolásának megszervezése az elemzéshez megfelelő formátumokban (Apache Parquet, CSV, ORC) (8.1-2. ábra).
- Kezdje el alkalmazni az alapvető statisztikai módszereket és dolgozzon reprezentatív mintákkal - vagy legalábbis ismerkedjen meg a statisztika alapelveivel (9.2-5. ábra).
- Az adatok és az adatok közötti kapcsolatok vizualizálására szolgáló eszközök alkalmazása és elsajátítása az elemzés eredményeinek vizualizálása érdekében. Jó vizualizáció nélkül lehetetlen magát az adatot vagy az azon alapuló folyamatokat teljes mértékben megérteni (7.1-4. ábra).

A nyílt adatformátumokra való áttérés és az információk elemzésére, tárolására és megjelenítésére szolgáló eszközök bevezetése megteremti a fenntartható és független digitális kormányzás alapjait. Ez nemcsak a szállítótól való függőséget csökkenti, hanem az adatokhoz való egyenlő hozzáférést is biztosítja valamennyi érdekelt fél számára.

8. lépés: Kezdje el a gépi tanulás végrehajtását az előrejelzéshez

Sok vállalat hatalmas mennyiségű adatot halmozott fel - egyfajta "információs gejzírként", amely még mindig kihasználatlanul áll. Ezeket az adatokat több száz vagy ezer projekt során gyűjtötték össze, de gyakran csak egyszer vagy egyáltalán nem használták fel a további folyamatokban. A zárt formátumokban és rendszerekben tárolt dokumentumokat és modelleket gyakran elavult és haszontalan ballasznak tekintik. A valóságban azonban értékes erőforrást jelentenek - az elkövetett hibák elemzésének, a rutinműveletek automatizálásának és a jövőbeli projektek automatikus osztályozására és a jellemzők felismerésére szolgáló innovatív megoldások kifejlesztésének alapját.

A fő kihívás az, hogy megtanuljuk, hogyan lehet ezeket az adatokat kinyerni és használható felismerésekkel alakítani. Amint azt a Gépi tanulás és előrejelzések című fejezetben tárgyaltuk, a gépi tanulási technikák jelentősen javíthatják a becslések és előrejelzések pontosságát számos építéssel kapcsolatos folyamatban. A felhalmozott adatok teljes körű felhasználása megnyitja az utat a hatékonyság javítása, a kockázatok csökkentése és a fenntartható digitális folyamatok kiépítése előtt.

Fő tevékenységek:

- Kezdje egyszerű algoritmusokkal:
 - Próbálja ki a lineáris regresszió alkalmazását - az LLM útmutatásainak felhasználásával - a visszatérő teljesítmény előrejelzésére olyan adathalmazokban, ahol a nagyszámú tényezőtől való függőség hiányzik vagy minimális (9.3-4. ábra).
 - Fontolja meg, hogy folyamatainak mely szakaszaiban lehetne elméletileg alkalmazni a k közelebbi szomszédok (k-NN) algoritmust - például osztályozási feladatoknál, tárgyak hasonlóságának értékelésénél vagy történelmi analógiák alapján történő előrejelzésnél (9.3-5. ábra).
- Gyűjtse és strukturálja az adatokat a modellek képzéséhez:
 - Gyűjtse össze a korábbi projektadatokat egy helyen és egységes formátumban (9.1-10. ábra).
 - A képzési minták minőségének és reprezentativitásának javítása az automatizált ETL (9.2-8. ábra) segítségével.
 - Tanuljuk meg az adatokat gyakorló és tesztkészletekre szétválasztani, ahogyan azt a Titanic adathalmaz példájánál tettük (9.2-6. ábra, 9.2-7. ábra).
- Fontolja meg a gépi tanulási technikák alkalmazásának kiterjesztését a feladatok széles körére, a projektidőzítések előrejelzésétől a logisztika optimalizálásig, az erőforrás-gazdálkodásig és a potenciális problémák korai felismeréséig.

A gépi tanulás egy olyan eszköz, amely az archivált adatokat értékes eszközzé alakítja előrejelzés, optimalizálás és megalapozott döntéshozatal céljából. Kezdje kis adathalmazokkal (9.2-5. ábra) és egyszerű modellekkel, majd fokozatosan növelje a komplexitást.

9. lépés: Integrálja az IoT és a modern adatgyűjtési technológiákat.

Az építőipar világa gyorsan digitalizálódik: minden építési fotó, minden csapatposzt már része a valóság paraméterezésének és tokenizálásának egy nagyobb folyamatának. Ahogyan a GPS egykor átalakította a logisztikát, az IoT, az RFID és az automatikus adatgyűjtési technológiák most az építőipart változtatják meg. Amint azt az "IoT Internet of Things és intelligens szerződések" című fejezetben tárgyaltuk, az érzékelőkkel és automatikus megfigyeléssel ellátott digitális építkezés jelenti az iparág jövőjét.

Fő tevékenységek:

- IoT -eszközök, RFID -címek bevezetése és a hozzájuk kapcsolódó folyamatok részletezése:
 - Annak értékelése, hogy a projekt mely területein vagy projektfázisaiban érhető el a legnagyobb megtérülés (ROI) - például a hőmérséklet, a rezgés, a páratartalom vagy a mozgásának nyomon követése esetén.
 - Fontolja meg az RFID használatát az anyagok, szerszámok és berendezések nyomon követésére az ellátási láncban.
 - Fontolja meg, hogy az összegyűjtött adatok hogyan integrálhatók egyetlen információs rendszerbe, például az Apache NiFi rendszerbe, automatizált, valós idejű feldolgozás és elemzés céljából (7.4-5. ábra).

■ Valós idejű felügyeleti rendszer létrehozása:

- A legfontosabb folyamat- vagy projektmetrikák nyomon követésére szolgáló műszerfalak fejlesztése olyan vizualizációs eszközökkel, mint a Streamlit, Flask vagy Power BI)
- Automatikus értesítések beállítása a tervtől vagy normáktól való kritikus eltérések jelzésére (7.4-2 ábra).
- Az összegyűjtött adatok és az azonosított minták alapján értékelje a berendezések megelőző karbantartásának lehetőségét (9.3-6. ábra).

■ Különböző forrásokból származó adatok kombinálása:

- Kezdje az adatmodell fizikai szintű vizualizálásával - tükrözze az információáramlás szerkezetét és a CAD-rendszerekből, IoT-eszközökből és ERP-platformokból származó kulcsfontosságú paramétereket (4.3-1 ábra).
- Kezdje az adatalemzés és a vezetői döntéstámogatás egységes platformjának durva vázlatának megalkotásával. Rögzítse a legfontosabb funkciókat, adatforrásokat, felhasználókat és a várható alkalmazási forgatókönyveket (4.3-7. ábra).

Minél hamarabb kezdi el összekapcsolni a valós folyamatokat a digitális világgal, annál hamarabb tudja azokat adatokkal kezelní - hatékonyan, átláthatóan és valós időben.

10. lépés. Készüljön fel az iparágban bekövetkező változásokra a jövőben

Az építőipari vállalatok folyamatosan a külső környezet nyomásának vannak kitéve: gazdasági válságok, technológiai ugrások, szabályozási változások. Mint egy erdő, amelynek ellen kell állnia esőnek, hónak, szárazságnak és tűzű napnak, a vállalatok is a folyamatos alkalmazkodás feltételei között élnek. És ahogy a fák is mély gyökérzetük révén válnak ellenállóvá a fagy és a szárazság ellen, úgy csak azok a szervezetek maradnak életképesek és versenyképesek, amelyek szilárd alapokon nyugvó automatizált folyamatokkal, a változások előrejelzésének képességével és rugalmasan alkalmazkodó stratégiákkal rendelkeznek.

Amint azt a "Túlélési stratégiák: versenyképes előnyök kiépítése" című fejezetben említettük, az építőipar a radikális átalakulás szakaszába lép. Az ügyfél és a vállalkozó közötti interakció a meggyőzés modellje felé halad, ahol az átláthatóság, a kiszámíthatóság és a digitális eszközök felváltják a hagyományos megközelítéseket. Ebben az új valóságban nem a legnagyobbak, hanem a legrugalmasabbak és technológiailag legfejlettebbek nyernek.

Fő tevékenységek:

■ Az üzleti sebezhetségek elemzése a nyílt adatok összefüggésében:

- Mérje fel, hogy az adatokhoz való hozzáférés demokratizálása az Uberizáció részeként milyen pusztító hatással lehet az Ön versenyzőjére és vállalkozására (10.1-5. ábra).
- Gondolkodjon el egy olyan stratégián, amely az átláthatatlan és silázott folyamatokról a nyílt megoldásokon, a rendszerek átvállatosságán és az adatok átláthatóságán alapuló üzleti modellek felé mozdul el (2.2-5. ábra).

■ Hosszú távú digitális stratégia kidolgozása:

- Határozza meg, hogy innovációs vezető kíván-e lenni, vagy inkább a "felzárkózási" forgatókönyvet részesíti előnyben, amelyben takarékoskodik az erőforrásaival.

- A szakaszok leírása: rövid távú (folyamatok automatizálása, adatstruktúrálás), középtávú (LLM és ETL megvalósítása), hosszú távú (digitális ökoszisztémák, központosított adattárak).
- Gondolkodjon szolgáltatási portfóliójának bővítésén:
 - Új szolgáltatások nyújtásának megfontolása (energiahatékonyságra, ESG-re, adatszolgáltatásokra összpontosítva). Az új üzleti modellek ről bővebben a következő fejezetben fogunk beszélni.
 - Törekedjen arra, hogy megbízható technológiai partnerként pozicionálja magát, amely támogatja a létesítmény teljes életciklusát a tervezéstől az üzemeltetésig. Az Önbe vetett bizalomnak a szisztematikus megközelítésen, az átlátható folyamatokon és a fenntartható technológiai megoldások nyújtásának képességén kell alapulnia.

Egy átalakuló környezetben nem azok nyernek, akik egyszerűen csak reagálnak a változásokra, hanem azok, akik proaktívak. A rugalmasság, a nyitottság és a digitális érettség a fenntarthatóság alapjai a holnap építésében.

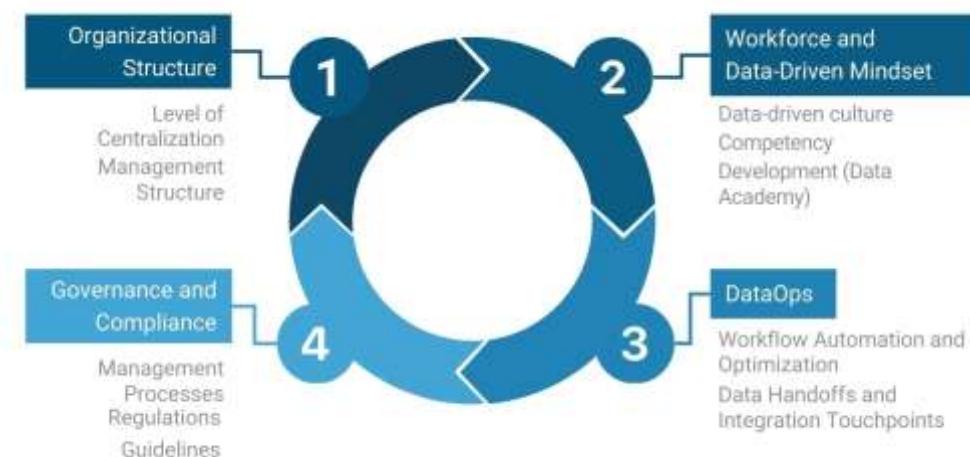
Átalakítási ütemterv: a káosztól az adatvezérelt vállalatig

Az alábbi terv kezdeti viszonyítási alapként szolgálhat - kiindulópontként saját adatvezérelt digitális átalakítási stratégiájának kialakításához:

- **Audit és szabványok:** a jelenlegi állapot elemzése, az adatok egységesítése
- **Adatok strukturálása és osztályozása:** a strukturálatlan és lazán strukturált adatok átalakításának automatizálása.
- **Csoportosítások, számítások és számítások automatizálása:** nyílt forráskódú eszközök és könyvtárak használata az automatizáláshoz.
- **Ökoszisztéma és COE:** belső csapat létrehozása a vállalaton belüli egységes adatökoszisztéma kialakítására.
- **Kultúra és tanulás:** elmozdulás a HiPPO -megoldásoktól az adatvezérelt megoldások felé.
- **Kísérletek, visszajelzés és skálázás:** járjon el iteratív módon: tesztelje az új módszereket korlátozott mértékben, gyűjtsön érvényes visszajelzéseket, és fokozatosan növelje a megoldások skáláját.
- **Nyílt formátumok:** univerzális és nyílt formátumok használata a szoftvergyártóktól való függetlenség érdekében.
- **Gépi tanulás:** az ML algoritmusok beágyazása a folyamatokba előrejelzés és optimalizálás céljából.
- **IoT és a digitális építkezés:** a modern adatgyűjtési technológiák integrálása a folyamatokba
- **Stratégiai alkalmazkodás:** felkészülés a jövőbeli ipari változásokra

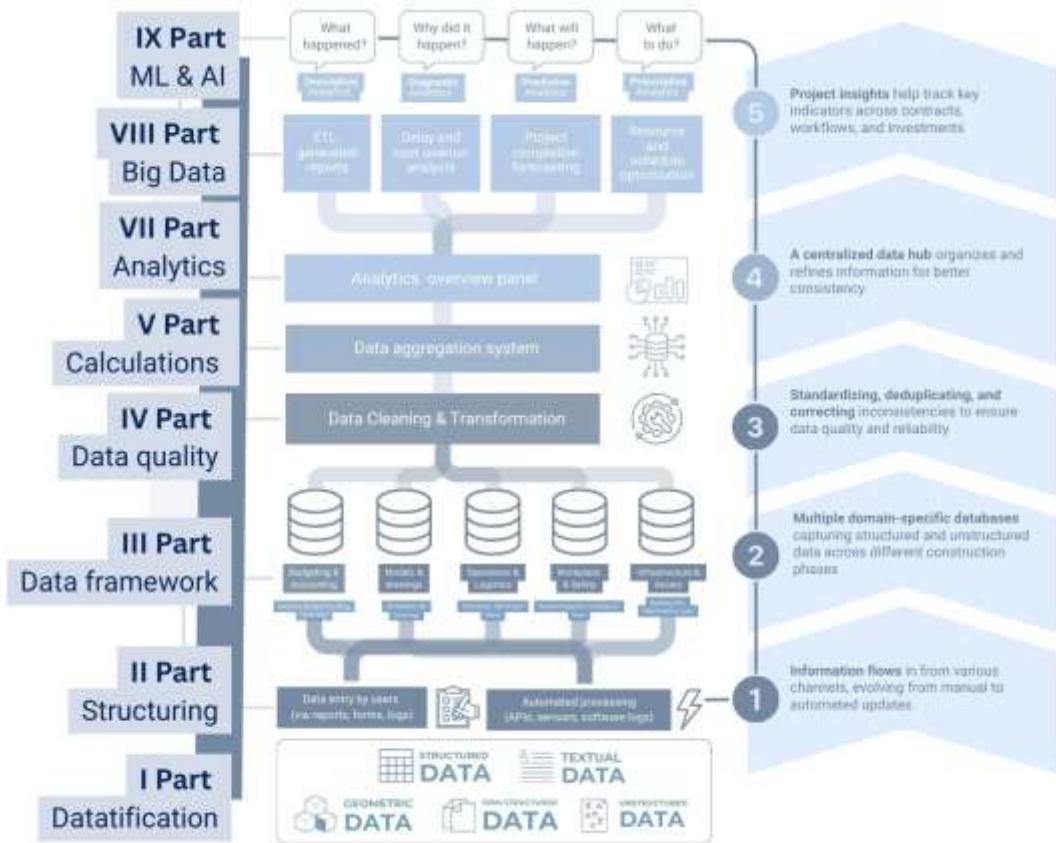
A legfontosabb, hogy ne feledje, hogy "az adatok önmagukban nem változtatják meg a vállalatot: azok az emberek változtatják meg, akik tudják, hogyan kell dolgozni az adatokkal". Hangsúlyozza a kultúrát, az átlátható folyamatokat és a folyamatos fejlesztés iránti elkötelezettséget.

A rendszerszemléletű megközelítés lehetővé teszi, hogy a digitális kezdeményezéseket a silózott digitális kezdeményezésekkel egy teljes körű adatvezérelt irányítási modellre váltsuk, ahol a döntések nem intuíciókon vagy feltételezésekben, hanem adatokon, tényeken és matematikailag kiszámított valószínűségeken alapulnak. Az építőipar digitális átalakulása nem csupán a technológia átvételéről szól, hanem egy olyan üzleti ökoszisztemára kialakításáról, ahol a projektinformációk zökkenőmentesen és iteratív módon kerülnek át a különböző rendszerek között. Ennek során a gépi tanulási algoritmusok automatikus, folyamatos elemzést, előrejelzést és folyamatoptimalizált biztosítanak. Egy ilyen környezetben a spekuláció és a rejtett adatok irrelevánssá válnak - csak a bevált modellek, az átlátható számítások és a kiszámítható eredmények maradnak.



10.2-3 A sikeres vállalati szintű adatkezelés legfontosabb elemei.

A könyv minden egyes része az építési projektek adatfeldolgozásának és elemzésének egy adott szakaszának felel meg (2.2-5. ábra). Ha vissza szeretne térti a korábban tárgyalt témák valamelyikéhez, és az adatfelhasználási folyamat holisztaikus megértése szempontjából szeretné áttekinteni, akkor a 10.2-4. ábrán található részcímekre hivatkozhat.



10.2-4. ábra A könyv részei az adatfeldolgozási csővezeték kontextusában (2.2-5. ábra): az információ digitalizálásától az analitikáig és a mesterséges intelligenciáig.

Függetlenül a szervezet méretétől, technológiai fejlettségi szintjétől vagy költségvetésétől, már ma megkezdheti az adatvezérelt megközelítés felé való elmozdulást. A helyes irányba tett apró lépések is idővel eredményeket hoznak.

Az adatvezérelt átalakítás nem egyszeri projekt, hanem folyamatos, iteratív fejlesztési folyamat, amely magában foglalja az új eszközök bevezetését, a folyamatok újratervezését és az adatvezérelt döntéshozatal kultúrájának kialakítását.

Építkezés az ipar 5.0-ban: Hogyan lehet pénzt keresni, ha már nem lehet elbújni

Az építőipari cégek sokáig a folyamatok átláthatatlanságán kerestek pénzt. A fő üzleti modell a spekuláció volt - az anyagköltségek, a munkamennyiség és a százalékos felárak túlbecsülése a zárt ERP - és PMIS - rendszerekben, amelyek külső ellenőrzések számára hozzáférhetetlenek. Az ügyfelek és megbízottjaik korlátozott hozzáférése az eredeti projektadatokhoz olyan rendszerekhez teremtette meg a terepet, amelyekben szinte lehetetlenné vált a számítások megbízhatóságának ellenőrzése.

Ez a modell azonban gyorsan veszít jelentőségeből. Az adathozzáférés demokratizálódásával, az LLM-ek megjelenésével, a nyílt adatok és az ETL automatizálási eszközök megjelenésével az iparág egy új munkastandard felé mozdul el.

Ennek eredményeképpen az átláthatatlanság már nem jelent versenyelőnyt - hamarosan olyan teherré válik, amelytől nehéz lesz megválni. Az átláthatóság opciós lehetőségből a piac maradás előfeltételévé válik.

Kikkel fognak az ügyfelek - bankok, befektetők, fizikai ügyfelek, magántőke, kormányzati ügyfelek - együtt dolgozni az új digitális valóságban? A válasz nyilvánvaló: azokkal, akik nemcsak az eredményt, hanem az ahhoz vezető út minden egyes lépésének indokoltságát is biztosítani tudják. A nyílt adatok mennyiségeinek növekedésével a partnerek és az ügyfelek olyan vállalatokat fognak választani, amelyek garantálják az eredmények átláthatóságát, pontosságát és kiszámíthatóságát.

Ennek fényében új üzleti modellek jelennek meg, amelyek a spekuláció helyett az adatkezelésen és a bizalmon alapulnak:

- **Négyzetméterek helyett folyamatok értékesítése:** a legfontosabb érték a bizalom és a hatékonyság lesz a konkrét megállapodások helyett. A fő érték a megbízható és ellenőrzött adatokon alapuló eredmény kiszámíthatósága lesz. A modern cégek nem az építési tárgyat mint olyat fogják eladni, hanem:
 - pontos határidők és átlátható munkarend;
 - ésszerű becslések, számításokkal alátámasztva;
 - teljes digitális nyomon követhetőség és ellenőrzés a projekt minden szakaszában.
- **Mérnöki és elemzési szolgáltatásként:** az "adat mint szolgáltatás" modell (kész adatok szolgáltatásként történő szolgáltatása a felhasználóknak az interneten keresztül), ahol minden projekt egy digitális adattovábbítási lánc részévé válik, és az üzleti értéket az adja, hogy ezt a láncot kezelní tudjuk. A vállalatok intelligens platformokká alakulnak át, amelyek automatizáláson és analitikán alapuló megoldásokat kínálnak:
 - A becslések és tervezések automatizált és átlátható elkészítése;
 - kockázat- és időzítésértékelés gépi tanulási algoritmusok alapján;
 - környezetvédelmi mutatók (ESG, CO₂, energiahatékonyság) kiszámítása;
 - jelentések készítése auditált nyílt forrásokból.
- **A mérnöki szaktudás produktivizálása:** a vállalat fejlesztései a vállalaton belül ismételten felhasználhatók és külön termékként forgalmazhatók - a digitális szolgáltatások révén további

bevételi forrást képezve. Az új környezetben a vállalatok nemcsak projekteket, hanem digitális eszközöket is létrehoznak:

- komponensek és becslési sablonok könyvtárai;
- automatizált ellenőrzési modulok;
- Nyílt forráskódú bővítmények és szkriptek (tanácsadás értékesítése) az adatokkal való munkához.

■ **Egy új vállalattípus: az adatvezérelt integrátor:** olyan piaci szereplő, amely nem függ bizonyos szoftvergyártótól vagy moduláris rendszerektől, és nincs egyetlen szoftverfelülethez "bezárva". Szabadon működik az adatokkal - és versenyképességét ezekre építi. A jövő építőipari vállalata nem csupán egy kivitelező, hanem egy olyan információs integrátor, amely a következő funkciókat képes ellátni a megrendelő számára:

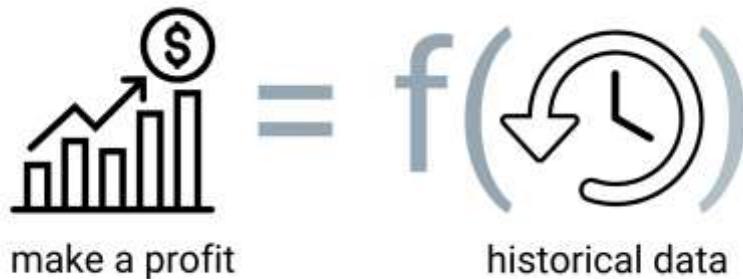
- Egyesítse a különböző forrásokból származó adatokat és végezzen elemzést;
- A folyamatok átláthatóságának és hitelességének biztosítása;
- tanácsadás az üzleti folyamatok optimalizálásával kapcsolatban;
- olyan eszközök fejlesztése, amelyek a nyílt adatok ökoszisztémájában működnek, LLM, ETL és Pipelines.

Az Ipar 5.0 (2.1-12. ábra) a "kézi átlagolási arányok korszakának" és a vezérigazgatók és a becslési és számviteli osztály közötti esti megbeszéléseknek a végét jelenti. minden, ami korábban rejte volt - számítások, becslések, mennyiségek - nyitottá, ellenőrizhetővé és érhetővé válik még a nem szakemberek számára is. Azok lesznek a nyertesek, akik elsőként tájékozódnak újra. mindenki más kimarad az építőipari ágazat új digitális gazdaságából.

KÖVETKEZTETÉS

Az építőipar az alapvető változások korszakába lép. Az első agyagtáblákon rögzített feljegyzésekkel kezdve a projekt-szerverekről és az építkezésekkel érkező hatalmas mennyiségű digitális adatig az építőipari információkezelés története mindig is a kor technológiájának érettsgégét tükrözte. Ma, az automatizálás, a nyílt formátumok és az intelligens elemzőrendszerek megjelenésével az ágazat nem fokozatos fejlődés, hanem gyors digitális átalakulás előtt áll.

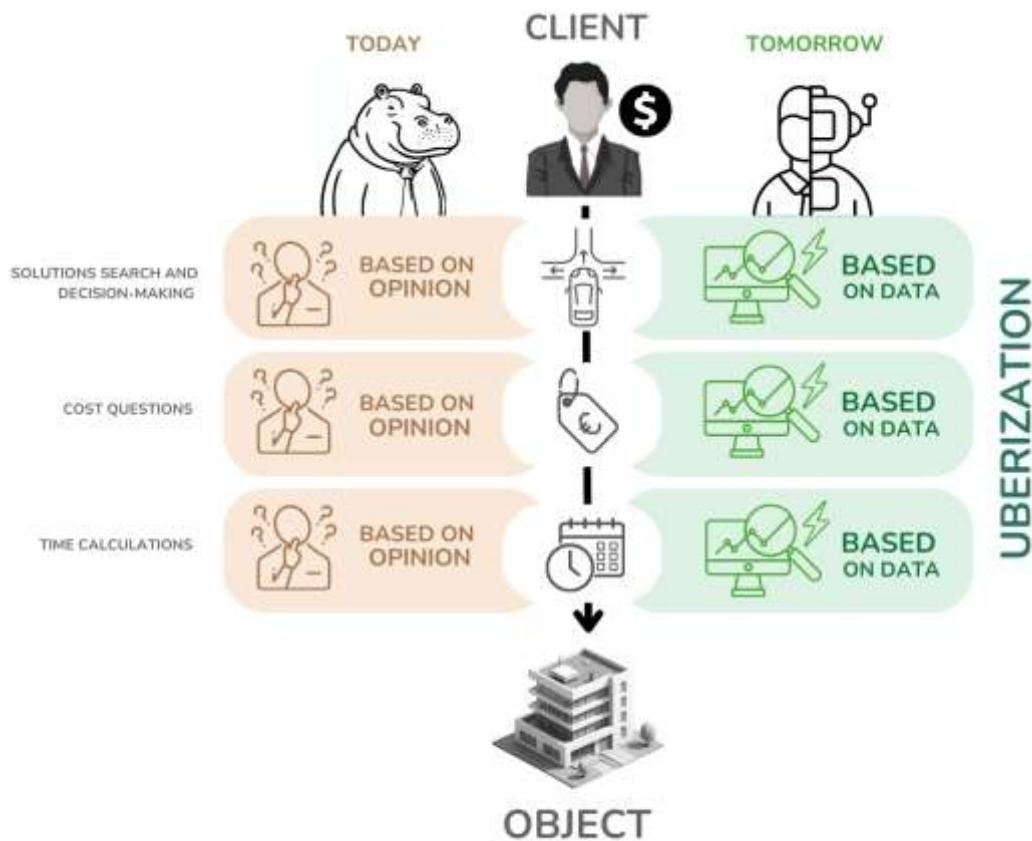
A gazdaság más ágazataihoz hasonlóan az építőiparban is újra kell gondolni nemcsak az eszközöket, hanem a munka elveit is. A korábban a piacot diktáló és az ügyfél és a projekt közötti fő közvetítő szerepet betöltő vállalatok elveszítik egyedülálló pozíciójukat. A bizalom és az adatokkal való munkavégzés képessége - az adatok gyűjtésétől és strukturálásától az elemzésig, előrejelzésig és a döntések automatizálásáig - előtérbe kerül.



10.2-1. ábra A strukturált múltbeli adatok a hatékony és irányítható vállalkozás üzemanyaga.

Ez a könyv részletesen ismerteti az építőipari adatkezelés legfontosabb alapelveit - az auditálástól és szabványosítástól kezdve a folyamatok automatizálásáig, a vizualizációs eszközök használatáig és az intelligens algoritmusok alkalmazásáig. Megvizsgáltuk, hogyan lehet még korlátozott erőforrásokkal is működő adatarchitektúrát kiépíteni, és hogyan lehet elkezdeni a döntéseket nem intuíciókra, hanem ellenőrizhető tényekre alapozva meghozni. Az adatokkal való munka már nem csupán az informatikai részleg feladata - az adatokkal való munka a vezetési kultúra alapjává válik, amely meghatározza a vállalat rugalmasságát, alkalmazkodóképességét és hosszú távú fenntarthatóságát.

A gépi tanulási technológiák, az automatikus feldolgozó rendszerek, a digitális ikrek és a nyílt formátumok alkalmazása már ma is lehetővé teszi az emberi tényező kiküszöbölését ott, ahol az korábban kritikus volt. Az építőipar az autonómia és az irányíthatóság felé halad, ahol az ötlettől a projekt megvalósításáig tartó folyamat a robotpilóta üzemmódban történő navigációhoz hasonlítható: szubjektív döntésekkel való függés nélkül, minden szakaszban kézi beavatkozás nélkül, de teljes digitális nyomon követhetőséggel és ellenőrzéssel (10.2-2. ábra).



Ábra 10.2-2 A fontos szakértők véleményén alapuló döntéshozatalról (HiPPO) az adatelemzésre való áttérést elsősorban az ügyfél fogja elősegíteni.

A könyvben bemutatott módszerek, elvek és eszközök elsajátításával Ön is képes lesz arra, hogy vállalatában ne intuitív, hanem adatvezérelt döntéseket hozzon. Képes lesz továbbá modulláncokat futtatni az LLM-ekben, a kész ETL-csővezetékeket bemásolni a fejlesztőkörnyezetébe (IDE), és automatikusan feldolgozni az adatokat, hogy a kívánt formában kapja meg a szükséges információkat. A könyv nagy adatokról és gépi tanulásról szóló fejezeteire építve ezután képes lesz összetettebb forgatókönyvek megvalósítására - új meglátások kinyerésére a múltbeli adatokból és gépi tanulási algoritmusok alkalmazására a folyamatok előrejelzésére és optimalizálására.

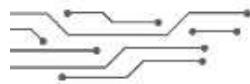
A nyílt adatok és folyamatok megalapozzák a projektköltségek és határidők pontosabb becslését, és megakadályozzák, hogy az építőipari vállalatok átláthatatlan adatokon spekuláljanak. Ez egyszerre jelent kihívást és lehetőséget az iparág számára, hogy újragondolja szerepét, és alkalmazkodjon egy olyan új környezethez, amelyben az átláthatóság és a hatékonyúság kulcsfontosságú sikertényezőkké válnak.

A digitális átalakulás korában a siker kulcsa a tudás átvételére és gyakorlati alkalmazására való hajlandóság.

Az új digitális versenyben azok a vállalatok lesznek előnyben, amelyek ezt előbb felismerik. Fontos azonban felismerni, hogy az adatok önmagukban nem változtatnak meg semmit. Sok embernek meg kell változtatnia a gondolkodásmódját, és ehhez ösztönzésre van szükség. A vállalatnak újra kell gondolnia az adatmegosztás módját.

Azok az emberek változtatják meg a vállalatot, akik tudják, hogyan kell ezekkel az adatokkal dolgozni, értelmezni, optimalizálni és új folyamatarchitektúrát létrehozni ezek alapján.

Ha ezeket a sorokat olvasod, akkor készen állsz a változásra, és már egy lépéssel előrébb jársz. Köszönöm, hogy ezt az utat választottad. Üdvözöljük a digitális átalakulás korszakában!



A SZERZŐRŐL

A nevem Artyom Boiko. Az építkezéseken való munkám 2007-ben kezdődött - bányászként dolgoztam egy olajpala-bányában, szülővárosomban, miközben a Szentpétervári Bányászati Egyetemen tanultam, bánya- és földalatti építés szakirányon. A könyv hátsó borítóján egy robbantó látható, ahol több száz kocka olajpala bányát bányásztunk és robbantottunk. Pályafutásom sokféle irányba vezetett, a bánya- és földalatti építőmunkástól kezdve az ipari mászó, tetőfedő és felvonószerelő munkásságon át. Abban a megtiszteltetésben volt részem, hogy különböző léptékű projektekben vehettem részt, a magánházak építésétől a nagy ipari létesítményekig a világ különböző régióiban.



Idővel munkám a fizikai építészről az információkezelésre és a digitális folyamatokra helyeződött át. 2013 óta különböző pozíciókban dolgoztam kis, közepes és nagy építőipari vállalatoknál Németország több régiójában, a tervezőtől az adatkezelési vezetőig. Ami az adatkezelést illeti, tapasztalataim különböző ERP-rendszerben, CAD (BIM), MEP, FEM, CMS adatokkal való munkából állnak. Foglalkoztam optimalizálással, folyamatautomatizálással, valamint elemzéssel, gépi tanulással, adatfeldolgozással az építési munkák tervezési, számítási és kivitelezési fázisaiban ipari, lakóépület-, infrastruktúra- és közműépítő vállalatoknál.

2003 óta dolgozom nyílt forráskódú szoftverekkel és nyílt adatokkal. Ez idő alatt számos webes projektet valósítottam meg - a weboldalaktól és online boltoktól kezdve a teljes értékű webes alkalmazásokig - nyílt forráskódú megoldások és nyílt CMS használatával. Ezek a platformok, amelyek sok tekintetben hasonlítanak a modern építőipari ERP-khez, moduláris felépítésűek, nagyfokú alkalmazkodóképességgel és hozzáférhetőséggel rendelkeznek. Ez a tapasztalat megalapozta szakmai megközelítésemet - a nyílt forráskódú technológiákra és a kollaboratív fejlesztés kultúrájára való összpontosítást. A nyílt forráskód és a szabad tudáscsere iránti tiszteletet igyekszem előmozdítani az építőiparban. Az építőiparban az adatok hozzáférhetőségének javítása érdekében végzett munkám több közösségi média közösség létrehozásában nyilvánult meg, amelyek az adatok nyitottságát és a nyílt forráskód építőipari alkalmazását vitatják meg, valamint több olyan startup vállalkozás elindításában, amelyek különböző zárt rendszerekből és platformokból származó adatokhoz való hozzáférést biztosító megoldásokat fejlesztenek.

A szakmai közösséghez való hozzájárulásom a CAD-interoperabilitás (BIM), ERP, 4D-5D, LLM Machine Learning és Artificial Intelligence témaúj konferenciákon való előadóként való részvétellel, valamint az európai építőipari kiadványokban megjelent cikkeimmel. Egyik kiemelkedő eredményem a "BIM History" létrehozása. [111], amely az építőiparban alkalmazott fontos adatkezelési szoftvermegoldások átfogó térképe. A több nyelvre lefordított, 7 részes "BIM-fejlesztés és lobbijátékok" című cikksorozatomat széles körben elismerték, mint a digitális szabványok rejtett dinamikájának megvilágítására tett kísérletet.

Így jutottam el a kőzetbányászattól az építési adatok bányászatáig és rendszerezéséig. Mindig nyitott vagyok a szakmai párbeszédre, új ötletekre és közös projektekre. Hálásan fogadok minden visszajelzést, és várom az üzeneteket, vagy várom, hogy a közösségi média követőim között találkozzunk. Köszönöm szépen, hogy végigolvastad ezt a könyvet! Örülök, ha ez a könyv segítene jobban megérteni az adatok témaát az építőiparban.

TOVÁBBI KAPCSOLAT

Az olvasói vélemények fontos szerepet játszanak a kiadványok továbbfejlesztésében és a kiemelt téma kiválasztásában. Különösen értékesek azok az észrevételek, amelyek arra vonatkoznak, hogy mely ötletek bizonyultak hasznosnak, és melyek azok, amelyek kétségeket vetettek fel, és további pontosítást vagy forrásmegjelölést igényelnek. A könyv sokféle anyagot és elemzést tartalmaz, amelyek közül néhány ellentmondásosnak vagy szubjektíven tűnhet. Ha az olvasás során pontatlanságokat, helytelenül idézett forrásokat, logikai ellentmondásokat vagy elgépeléseket talál, hálás lennék az észrevételeiről, gondolatairól vagy kritikáiról, amelyeket a következő címre küldhet: boikoartem@gmail.com. Vagy üzeneteken keresztül a LinkedIn-en: linkedin.com/in/boikoartem.

Örülök, ha a közösségi médiában megemlítenék az Adatvezérelt építés című könyvet - az olvasási tapasztalatok megosztása segít a nyílt adatok és eszközök terjesztésében, és támogatja a munkámat.

FORDÍTÁS KOMMENTÁR

Ezt a könyvet mesterséges intelligencia technológiával fordították le. Ez sokkal gyorsabbá tette a fordítási folyamatot. Azonban mint minden technológiai műveletnél, itt is előfordulhatnak hibák vagy pontatlanságok. Ha bármilyen hibásnak vagy helytelenül lefordítottnak tűnő dolgot észlel, kérjük, küldjön nekem e-mailt. Észrevételei segítenek a fordítás minőségének javításában.

ADATVEZÉRELT ÉPÍTŐIPARI KÖZÖSSÉGEK

Ez egy olyan hely, ahol szabadon felteheti kérdéseit, megoszthatja problémáit és megoldásait:

DataDrivenConstruction.io: <https://datadrivenconstruction.io>

LinkedIn: <https://www.linkedin.com/company/datadrivenconstruction/>

Twitter: <https://twitter.com/datadrivenconst>

Telegram: <https://t.me/datadrivenconstruction>

YouTube: <https://www.youtube.com/@datadrivenconstruction>

EGYÉB KÉSZSÉGEK ÉS FOGALMAK

Az építőiparban az adatokkal való munka legfontosabb alapelvei mellett a DataDrivenConstruction számos további fogalommal, programmal és készseggyel foglalkozik, amelyek elengedhetetlenek az adatvezérelt szakemberek számára. Ezek közül néhányat csak áttekintő jelleggel mutatunk be, de a gyakorlat szempontjából kritikus jelentőségűek.

Az érdeklődő olvasó a DataDrivenConstruction.io weboldalon további, a kulcskompetenciákkal kapcsolatos anyagokra mutató linkekkel talál. Ezek az anyagok tartalmazzák a Python és Pandas -val való munkát, az ETL -folyamatok felépítését, az építőipari CAD-projektek adatfeldolgozásának példáit, a nagy adatrendszereket, valamint az építőipari adatok vizualizációjának és analitikájának modern megközelítéseit.

Az "DataDrivenConstruction" című könyv és az összes esettanulmány elkészítéséhez számos nyílt forráskódú eszközt és szoftvert használtak fel. A szerző szeretne köszönetet mondani az alábbi megoldások fejlesztőinek és társszerzőinek:

- Python és Pandas - az adatmanipuláció és automatizálás gerince
- Scipy, NumPy, Matplotlib és Scikit-Learn - könyvtárak adatalemzéshez és gépi tanuláshoz
- SQL és Apache Parquet - nagy mennyiségű építési adat tárolására és feldolgozására szolgáló eszközök.
- Nyílt forráskódú CAD (BIM) nyílt formátumú, nyílt adatszolgáltatási eszközök
- N8n, Apache Airflow, Apache NiFi - orchestrációs és munkafolyamat-automatizálási rendszerek.
- DeepSeek, LLaMa, Mistral - Nyílt forráskódú LLM

Külön köszönet a szakmai közösségekben és a közösségi hálózatokban a nyílt adatok és eszközök témaáról folytatott viták minden résztvevőjének, akiknek kritikái, észrevételei és ötletei hozzájárultak a könyv tartalmának és szerkezetének javításához.

Kövesse a projekt fejlődését a DataDrivenConstruction.io weboldalon, ahol nemcsak a könyv frissítéseit és javításait, hanem új fejezeteket, útmutatókat és gyakorlati példákat is közzétesznek a leírt technikák alkalmazására.

MAXIMÁLIS KÉNYELEM A NYOMTATOTT VERZIÓVAL

Ön birtokában van a **Data-Driven Construction** ingyenes digitális verziója. A kényelmesebb munkavégzés és az anyagokhoz való gyors hozzáférés érdekében javasoljuk, hogy figyeljen a **nyomtatott kiadásra**:



■ **Mindig kéznél van:** a nyomtatott formátumú könyv megbízható munkaeszközzé válik, amely lehetővé teszi, hogy bármilyen munkaszituációban gyorsan megtalálja és használja a megfelelő ábrákat és diagramokat.

■ **Kiváló minőségű illusztrációk:** a nyomtatott kiadásban szereplő összes kép és grafika maximális minőségen jelenik meg.

■ **Gyors hozzáférés az információkhoz:** könnyű navigáció, jegyzetek, könyvjelzők készítése és a könyvvel való munka bárhol.

A könyv teljes nyomtatott változatának megvásárlásával egy kényelmes eszközt kap az információkkal való kényelmes és hatékony munkavégzéshez: a vizuális anyagok gyors használatának lehetőségét a minden nap feladatakban, a szükséges sémák gyors megtalálását és jegyzetelését. Emellett vásárlásával támogatja a nyílt tudás terjesztését.

Rendelje meg a könyv nyomtatott változatát a következő címen:

datadrivenconstruction.io/books



EGYEDÜLÁLLÓ LEHETŐSÉG A STRATÉGIAI POZÍCIONÁLÁS

Felajánljuk, hogy reklámanyagokat helyezzen el a DataDrivenConstruction ingyenes verziójában. A kiadvány fizetős változata a megjelenését követő első évben a világ több mint 50 országából - Latin-Amerikától az ázsiai-csendes-óceáni térségig - érkező szakemberek figyelmét keltette fel. Az együttműködés egyéni feltételeinek megbeszéléséhez és az elhelyezési lehetőségekről való részletes tájékoztatáshoz kérjük, töltse ki a datadrivenconstruction.io hivatalos portálon található visszajelző űrlapot, vagy írjon a könyv végén felsorolt elérhetőségekre.



A KÖNYV FEJEZETEI ELÉRHETŐEK A DATADRIVENCONSTRUCTION.IO OLDALON.

Az Adavezérelt építés fejezetait az Adavezérelt építés weboldalon olvashatja, ahol a könyv egyes részeit fokozatosan közzétesszük, hogy Ön gyorsan megtalálja a szükséges információkat, és felhasználhassa azokat munkája során. Emellett számos más, hasonló témaúj kiadványt, valamint alkalmazási példákat és megoldásokat is talál, amelyek segítik Önt képességeinek fejlesztésében és az adatok építőiparban való alkalmazásában.



A KÖNYV LEGÚJABB VÁLTOZATAI LETÖLTHETŐ A HIVATALOS HONLAPRÓL

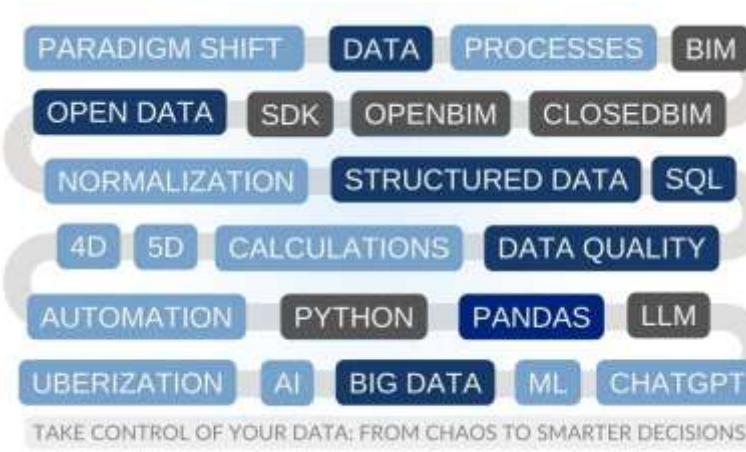
A DataDrivenConstruction könyv aktuális és legújabb verziója letölthető a datadrivenconstruction.io oldalról. Ha szeretne frissítéseket kapni a könyv új fejezeteiről, gyakorlati tippekről vagy új alkalmazásokról szóló ismertetőkről, iratkozzon fel a hírlevélre:

- Ön lesz az első, aki megismerekedik a könyv új részeivel.
- Gyakorlati esettanulmányok és tanácsok az építőiparban alkalmazott analitikáról és automatizálásról.
- Tartsa naprakészen magát a trendekkel, kiadványokkal és alkalmazási példákkal kapcsolatban.

Menj a datadrivenconstruction.io oldalra, hogy feliratkozz!

ADATVEZÉRELT ÉPÍTÉS: TANÁCSADÁS, WORKSHOPOK ÉS KÉPZÉS

A DataDrivenConstruction képzési programjai és tanácsadásai világszerte több tucat vezető építőipari vállalatnak segítettek a hatékonyság növelésében, a költségek csökkentésében és a megoldások minőségének javításában. A DataDrivenConstruction ügyfelei közé tartoznak a milliárd eurós piac legnagyobb szereplői, köztük építőipari, tanácsadó és informatikai vállalatok.



Miért válasszon minket?

- **Relevancia:** az iparág főbb trendjeiről és meglátásairól való beszélgetés.
- **Gyakorlati:** a szakemberek segítése a minden nap feladatok hatékony megoldásában a PoC segítségével.
- **Személyre szabott megközelítés:** az Ön vállalkozására szabva, a képzés és tanácsadás előnyeinek maximalizálása.

A DataDrivenConstruction csapat fő fókuszterületei:

- **Adatminőség-menedzsment:** segít a feladatok paraméterezésében, a követelmények összegyűjtésében, az adatok érvényesítésében és előkészítésében az automatizált feldolgozáshoz.
- **Adatbányászat - adatok kinyerése és strukturálása:** ETL folyamatok beállítása és adatok kinyerése e-mailkből, PDF, Excel, képekből és egyéb forrásokból.
- **BIM és CAD analitika:** összegyűjti, strukturálja és elemzi az RVT fájlok ból származó információkat, IFC, DWG és más CAD (BIM) formátumokból.
- **Elemzés és adatátalakítás:** a széttagolt információk strukturált adatokká, elemzésekkel, meglátásokká és döntésekkel alakítása.
- **Adatintegráció és folyamatautomatizálás:** az automatikus dokumentumkészítéstől a belső rendszerekkel és külső adatbázisokkal való integrációig.

Vegye fel a kapcsolatot a DataDrivenConstruction.io-val, hogy megtudja, hogyan segíthet az automatizálás használata vállalatának kézzelfogható üzleti eredmények elérésében.

FOGALOMTÁR

AI (Artificial Intelligence) - Mesterséges intelligencia; a számítógépes rendszerek képessége olyan feladatok elvégzésére, amelyekhez általában emberi intelligencia szükséges, mint például a mintafelismerés, a tanulás és a döntéshozatal.

Az Apache Airflow egy nyílt forráskódú munkafolyamat-rendezési platform, amely lehetővé teszi a munkafolyamatok és ETL-ek programozott létrehozását, ütemezését és nyomon követését DAG-ok (irányított aciklikus gráfok) segítségével.

Az Apache NiFi a rendszerek közötti adatáramlás automatizálására szolgáló eszköz, amely az adatok továbbítására és átalakítására specializálódott.

Az Apache Parquet egy hatékony fájlformátum oszlopos adattároláshoz, amelyet a nagy adatelemző rendszerekben való használatra optimalizáltak. Jelentős tömörítést és gyors feldolgozást biztosít.

API (Application Programming Interface) - egy formalizált interfész, amely lehetővé teszi, hogy az egyik program a forráskódhoz való hozzáférés nélkül kölcsönhatásba lépjen egy másik programmal, és szabványosított kéréseken és válaszokon keresztül adatokat és funkciókat cseréljen.

Attribútum - egy tárgy olyan jellemzője vagy tulajdonsága, amely leírja annak tulajdonságait (pl. terület, térfogat, költség, anyag).

Az adatbázisok az információk tárolására, kezelésére és elérésére szolgáló szervezett struktúrák, amelyeket az adatok hatékony visszakeresésére és feldolgozására használnak.

BEP (BIM végrehajtási terv) - Az épületinformációs modellezés megvalósítási terve, amely meghatározza a BIM projektben történő megvalósításának célkitűzéseit, módszereit és folyamatait.

Big Data - jelentős mennyiségű, változatos és frissítési sebességű információtömbök, amelyek feldolgozásához és elemzéséhez különleges technológiákra van szükség.

BI (Business Intelligence) - Üzleti intelligencia; olyan folyamatok, technológiák és eszközök, amelyekkel az adatokat a döntéshozatalhoz szükséges értelmes információkká alakíthatjuk.

BIM (Building Information Modeling) - Épületinformációs modellezés; az építési projektek fizikai és funkcionális jellemzőinek digitális ábrázolásának létrehozása és kezelése, amely nemcsak 3D modelleket, hanem a teljesítményre, az anyagokra, az időzítésre és a költségekre vonatkozó információkat is tartalmaz.

BlackBox/WhiteBox - a rendszer megértésének megközelítése: az első esetben a belső logika rejte van, csak a bemenetek és a kimenetek láthatók; a második esetben a feldolgozás átlátható és elemezhető.

A Bounding Box egy geometriai konstrukció, amely egy objektum határait írja le a háromdimenziós térben a minimális és maximális X, Y és Z koordináákkal, létrehozva egy "dobozt" az objektum körül.

A BREP (Boundary Representation) az objektumok geometriai ábrázolása, amely a felületek határain keresztül határozza meg azokat.

A CAD (Computer-Aided Design) egy számítógépes tervezési rendszer, amelyet pontos rajzok és 3D modellek létrehozására, szerkesztésére és elemzésére használnak az építészetben, az építőiparban, a mérnöki és egyéb iparágakban.

A CAFM (Computer-Aided Facility Management) egy olyan ingatlan- és infrastruktúra-menedzsment szoftver, amely magában foglalja a helyiségszervezést, az eszközgazdálkodást, a karbantartást és a költségek nyomon követését.

CDE (Common Data Environment) - egy központosított digitális tér a projektinformációk kezelésére, tárolására, megosztására és együttműködésére a létesítmény életciklusának minden szakaszában.

A kiválósági központ egy szervezeten belüli szakosodott struktúra, amely egy adott tudásterület fejlesztéséért, szabványok és legjobb gyakorlatok kidolgozásáért, a személyzet képzéséért és az innovációk bevezetésének támogatásáért felelős.

A CoClass egy modern, harmadik generációs épületelem-osztályozási rendszer.

A fogalmi adatmodell az alapvető entitások és kapcsolataik magas szintű, attribútum részletesség nélküli ábrázolása, amelyet az adatbázis-tervezés kezdeti szakaszában használnak.

A CRM (Customer Relationship Management) egy ügyfélkapcsolat-kezelő rendszer, amelyet az értékesítési és szolgáltatási folyamatok automatizálására használnak.

A DAG (Directed Acyclic Graph, irányított aciklikus gráf) egy irányított aciklikus gráf, amelyet az adatszervezési rendszerekben (Airflow, NiFi) a feladatsorozatok és függősségek meghatározására használnak.

A Dash egy Python keretrendszer interaktív webalapú adatvizualizációk létrehozására.

Dashboard - A **műszerfal**, amely vizuálisan, valós időben mutatja be a legfontosabb teljesítménymutatókat és mérőszámokat.

Az adatközpontú megközelítés egy olyan módszertan, amely az adatokat az alkalmazások vagy a szoftverkódok elő helyezi, és az adatokat a szervezet központi eszközévé teszi.

Adatkezelés - olyan gyakorlatok, folyamatok és irányelvek összessége, amelyek biztosítják az adatok megfelelő és hatékony felhasználását egy szervezeten belül, beleértve a hozzáférés, a minőség és a biztonság ellenőrzését.

Az adattó egy olyan tárolóhely, amelyet arra terveztek, hogy nagy mennyiségű nyers adatot eredeti formátumban tároljon a felhasználásig.

A Data Lakehouse egy olyan architektúrális megközelítés, amely egyesíti az adattavak (Data Lake) rugalmasságát és skálázhatóságát az adattárházak (DWH) kezelhetőségével és teljesítményével.

Az adatvezérelt építés olyan stratégiai megközelítés, amelyben a létesítmény életciklusának minden szakaszát - a tervezéstől az üzemeltetésig - automatizált, összekapcsolt rendszerek támogatják. Ez a megközelítés folyamatos, tényeken alapuló tanulást biztosít, csökkenti a bizonytalanságot, és lehetővé teszi a vállalatok számára a fenntartható iparági vezető szerep elérését.

Adatvezérelt integrátor - olyan vállalat, amely különböző forrásokból származó adatok kombinálására és elemzésére szakosodott, hogy vezetői döntéseket hozzon.

Adatvezérelt megközelítés - olyan módszertan, amelyben az adatokat stratégiai eszközök tekintik, és a döntéseket nem szubjektív vélemények, hanem az információk objektív elemzése alapján hozzák meg.

Adatminimalizmus - az adatoknak a legértékesebbre és legjelentősebbre való redukálására irányuló megközelítés, amely lehetővé teszi az információk egyszerűsített feldolgozását és elemzését.

Adatmocsár - A strukturálatlan adatok szétszórt tömege, amely akkor keletkezik, amikor az információkat megfelelő szervezés nélkül, ellenőrizetlen módon gyűjtik és tárolják.

A DataOps egy olyan módszertan, amely a DevOps elveket, az adatokat és az analitikát ötvözi, és az adatfolyamok együttműködésének, integrációjának és automatizálásának javítására összpontosít.

Az információ digitalizálása az építőipari tevékenység valamennyi aspektusának elemzésre, értelmezésre és automatizálásra alkalmas digitális formába történő átalakítása.

DataFrame - A Pandas könyvtár kétdimenziós táblázatos adatszerkezete, ahol a sorok az egyes rekordokat vagy objektumokat, az oszlopok pedig azok jellemzőit vagy attribútumait jelölik.

Leíró analitika - A múltbeli adatok elemzése annak megértése érdekében, hogy mi történt a múltban.

Diagnosztikai analitika - Az adatok elemzése annak megállapítására, hogy miért történt valami.

A Gantt-diagram egy olyan projekttervezési eszköz, amely a feladatokat vízszintes sávokként ábrázolja egy idővonalon, lehetővé téve a munka sorrendjének és időtartamának megjelenítését.

A DWH (Data Warehouse) egy olyan központosított adattárház-rendszer, amely több forrásból származó információkat aggregál, strukturál, és analitikai és jelentési célokra teszi elérhetővé.

ESG (Environmental, Social, Governance) - egy kritériumrendszer egy vállalat vagy projekt környezeti, társadalmi és irányítási hatásainak értékelésére.

Az ELT (Extract, Load, Transform) egy olyan folyamat, amelynek során az adatokat először kivonják a forrásokból és betöltenek egy adattárba, majd analitikai célokra átalakítják.

Az ETL (Extract, Transform, Load) az adatok különböző forrásokból történő kinyerése, a kívánt formátumba történő átalakítása és a céltárolóba történő betöltése elemzés céljából.

ER-diagram (Entity-Relationship) - az adatmodellezésben használt vizuális diagram, amely az entitásokat, azok attribútumait és a köztük lévő kapcsolatokat mutatja.

Az ERP (Enterprise Resource Planning) egy átfogó moduláris vállalati erőforrás-tervezési rendszer, amelyet az építési folyamat különböző aspektusainak kezelésére és optimalizálására használnak.

Jellemzők - A gépi tanulásban a modell bemeneteként használt független változók vagy attribútumok.

Fizikai adatmodell - az adatbázis szerkezetének részletes ábrázolása, beleértve a táblákat, oszlopokat, adattípusokat, kulcsokat és indexeket, egy adott DBMS-re optimalizálva.

Az FPDF egy Python könyvtár PDF dokumentumok létrehozására.

A Geometric Core egy olyan szoftverkomponens, amely alapvető algoritmusokat biztosít a CAD, BIM és más mérnöki alkalmazásokban a geometriai objektumok létrehozásához, szerkesztéséhez és elemzéséhez.

HiPPO (Highest Paid Person's Opinion) - a döntéshozatal olyan megközelítése, amely az objektív adatok helyett a szervezet legjobban fizetett személyének véleményén alapul.

IDE (Integrated Development Environment) - integrált fejlesztőkörnyezet, átfogó eszköz a kód írásához, teszteléséhez és hibakereséshez (pl. PyCharm, VS Code, Jupyter Notebook).

Az IDS (Information Delivery Specification) egy információs szolgáltatási specifikáció, amely meghatározza az adatigényeket a projekt különböző szakaszaiban.

Az IFC (Industry Foundation Classes) egy BIM adatcsere-formátum, amely biztosítja a különböző szoftvermegoldások közötti átjárhatóságot.

Az Ipar 5.0 egy olyan ipari fejlesztési koncepció, amely a digitalizáció, az automatizálás és a mesterséges intelligencia képességeit az emberi potenciállal és a környezeti fenntarthatósággal ötvözi.

Az adatintegráció a különböző forrásokból származó adatok egyetlen, koherens rendszerben történő egyesítése, amely egységes információszemléletet biztosít.

Az információs silók olyan elszigetelt adattároló rendszerek, amelyek nem osztják meg az információkat más rendszerekkel, akadályozva ezzel az adatok hatékony felhasználását.

Az IoT (Internet of Things) a fizikai tárgyak internethöz való csatlakoztatásának koncepciója az adatok gyűjtése, feldolgozása és továbbítása érdekében.

A k-NN (k-Nearest Neighbors) egy olyan gépi tanulási algoritmus, amely az objektumokat a gyakorló minta legközelebbi szomszédaihoz való hasonlóság alapján osztályozza.

A Kaggle egy adataelemzési és gépi tanulási versenyek platformja.

Kalkuláció - az építési munkák vagy folyamatok költségének kiszámítása egy bizonyos mértékegységre (pl. 1 m² gipszkartonfal, 1 m³ beton).

A KPI-k (Key Performance Indicators, azaz kulcsfontosságú teljesítménymutatók) olyan számszerűsíthető mérőszámok, amelyeket egy vállalat vagy egy adott projekt sikerének értékelésére használnak.

Címkek - A gépi tanulásban azok a célváltozók vagy attribútumok, amelyeket a modellnek meg kell jósolnia.

Tanulási algoritmus - Egy célfüggvénynek megfelelő modellben a legjobb hipotézis megtalálásának folyamata egy képzési adathalmaz felhasználásával.

Lineáris regresszió - Egy függő változó és egy vagy több független változó közötti kapcsolat modellezésére szolgáló statisztikai módszer.

LLM (Large Language Model) - Nagy nyelvi modell, egy mesterséges intelligencia, amelyet arra képezték ki, hogy hatalmas adatmennyiségekből szöveget értsen és generáljon, képes a szövegkörnyezet elemzésére és programozási kód írására.

LOD (Level of Detail/Development) - a modell részletességi szintje, amely meghatározza a geometriai pontosság és az információtartalom mértékét.

A logikai adatmodell az üzleti információkat és szabályokat tükröző entitások, attribútumok, kulcsok és kapcsolatok részletes leírása, amely a fogalmi és fizikai modellek közötti köztes szakasz.

Gépi tanulás - A mesterséges intelligencia technikák egy olyan osztálya, amely lehetővé teszi a számítógépes rendszerek számára, hogy explicit programozás nélkül tanuljanak és előrejelzéseket készítsenek adatokból.

A Masterformat egy első generációs osztályozási rendszer, amelyet az építési előírások szakaszokra és szakágakra történő strukturálására használnak.

MEP (Mechanical, Electrical, Plumbing) - épületgépészeti rendszerek, amelyek mechanikai, elektromos és vízvezeték-szerelési komponenseket tartalmaznak.

A háló a 3D objektumok hálós reprezentációja, amely csúcsokból, élekből és felületekből áll.

Modell - A gépi tanulásban különböző hipotézisek halmaza, amelyek közül az egyik megközelíti a megjósolandó vagy közelítendő célfüggvényt.

Az adatmodellezés az adatok és kapcsolataik strukturált reprezentációjának létrehozása az információs rendszerekben való megvalósítás céljából, beleértve a fogalmi, logikai és fizikai szinteket.

Az n8n egy nyílt forráskódú eszköz a munkafolyamatok automatizálására és az alkalmazások alacsony kódú megközelítéssel történő integrálására.

Normalizálás - a gépi tanulásban az a folyamat, amelynek során a különböző numerikus adatokat közös skálára hozzák a feldolgozás és elemzés megkönyítése érdekében.

Reverse engineering - egy tárgy eszközének, működésének és gyártási technológiájának tanulmányozása a tárgy szerkezetének, funkcióinak és működésének elemzésével. Az adatokkal összefüggésben - információk kinyerése védett formátumokból nyílt rendszerekben való felhasználás céljából.

Az OCR (Optical Character Recognition) egy **optikai** karakterfelismerő technológia, amely szöveges képeket (szkennelt dokumentumok, fényképek) alakít át gépileg olvasható szövegformátummá.

Az OmniClass egy második generációs nemzetközi osztályozási szabvány az építőipari információkezeléshez.

Ontológia - A fogalmak összefüggéseinek rendszere, amely formalizálja a tudás egy adott területét.

Nyílt forráskód - a szabad felhasználásra, tanulmányozásra és módosításra rendelkezésre álló nyílt forráskódú szoftverek fejlesztésének és terjesztésének modellje.

Az Open BIM a nyílt BIM koncepciója, amely magában foglalja a nyílt szabványok és formátumok használatát a különböző szoftvermegoldások közötti adatcseréhez.

Nyílt szabványok - nyilvánosan elérhető specifikációk egy adott feladat megvalósítására, amelyek lehetővé teszik a különböző rendszerek számára az interakciót és az adatcserét.

A Pandas egy nyílt forráskódú Python könyvtár adatfeldolgozáshoz és elemzéshez, amely DataFrame és Series adatstruktúrákat biztosít a táblázatos információk hatékony kezeléséhez.

A nyílt adatparadigma az adatfeldolgozás olyan megközelítése, amelyben az információkat szabadon hozzáférhetővé teszik, hogy bárki felhasználhassa, újrafelhasználhassa és terjeszthesse őket.

A parametrikus módszer egy olyan építési projektbecslési módszer, amely statisztikai modellekkel használ a projekt paramétereit alapján történő értékbecsléshez.

A PIMS (Project Information Model) egy digitális rendszer, amelyet a projektinformációk rendszerezésére, tárolására és megosztására terveztek.

Pipeline - Az adatfeldolgozási folyamatok sorozata, a kinyeréstől és átalakítástól az elemzésig és a megjelenítésig.

A PMIS (Project Information Management System) egy olyan projektirányítási rendszer, amelyet az egyes építési projektek szintjén történő részletes feladatellenőrzésre terveztek.

A prediktív analitika az analitikának egy olyan része, amely statisztikai módszereket és gépi tanulást használ a jövőbeli eredmények előrejelzésére a múltbeli adatok alapján.

Prescriptive Analytics - Az analitika egy olyan része, amely nemcsak a jövőbeli eredményeket jósolja meg, hanem optimális intézkedéseket is javasol a kívánt eredmények eléréséhez.

A védett formátumok olyan zárt adatformátumok, amelyeket egy adott vállalat ellenőriz, és amelyek korlátozzák az információ megosztásának lehetőségét, és növelik a függőséget az adott szoftvertől.

A QTO (Quantity Take-Off) az a folyamat, amelynek során a tervdokumentumokból kivonjuk az elemek mennyiségét, hogy kiszámíthassuk a projekt megvalósításához szükséges anyagmennyiségeket.

Minőségirányítási rendszer - minőségirányítási rendszer, amely biztosítja, hogy a folyamatok és az eredmények megfeleljenek a megállapított követelményeknek.

A RAG (Retrieval-Augmented Generation) egy olyan módszer, amely a nyelvi modellek generatív képességeit kombinálja a releváns információk kinyerésével a vállalati adatbázisokból, javítva a válaszok pontosságát és relevanciáját.

Az RDBMS (Relational Database Management System) egy relációs adatbázis-kezelő rendszer, amely az információkat egymáshoz kapcsolódó táblázatok formájában szervezi.

A RegEx (Regular Expressions) egy formalizált nyelv a karakterláncok keresésére és feldolgozására, amely lehetővé teszi, hogy sablonokat adjon meg a szöveges adatok bizonyos kritériumoknak való megfelelésének ellenőrzésére.

A regresszió a változók közötti kapcsolat elemzésének statisztikai módszere.

A CO₂-számítások az építőanyagok és eljárások előállításához és használatához kapcsolódó szén-dioxid-kibocsátás becslésére szolgáló módszer.

Az erőforrás-módszer az építési munkák elvégzéséhez szükséges összes erőforrás (anyagok, munkaerő, berendezések) részletes elemzésén alapuló becslési módszer.

Az RFID (rádiófrekvenciás azonosítás) egy olyan technológia, amely a tárgyak automatikus azonosítására szolgál rádiójelek segítségével, és amelyet anyagok, gépek és személyzet nyomon követésére használnak.

A ROI (Return on Investment) egy olyan mutató, amely a nyereség és a befektetett pénzeszközök közötti arányt tükrözi, és a befektetések hatékonyságának értékelésére szolgál.

A SaaS (Software as a Service - szoftver mint szolgáltatás) a szoftver mint szolgáltatás modellje, amelyben az alkalmazásokat egy szolgáltató hosztolja, és az interneten keresztül teszi elérhetővé a felhasználók számára.

SCM (Supply Chain Management) - ellátási lánc menedzsment, amely magában foglalja az összes folyamat koordinálását és optimalizálását az anyagbeszerzéstől a késztermékek leszállításáig.

Az adatsílok olyan elszigetelt információaktárak egy szervezeten belül, amelyek nincsenek integrálva más rendszerekkel, ami megnehezíti az adatmegosztást és nem hatékony.

Az SQL (Structured Query Language) egy strukturált lekérdezési nyelv, amelyet relációs adatbázisokkal való munkához használnak.

Az SQLite egy könnyű, beágyazható, platformokon átívelő DBMS, amely nem igényel külön szervert, és támogatja az alapvető SQL-funkciókat, széles körben használják mobil alkalmazásokban és beágyazott rendszerekben.

Strukturált adatok - meghatározott formátumban, egyértelmű struktúrával rendezett információk, például relációs adatbázisokban vagy táblázatokban.

Lazán strukturált adatok - részlegesen szervezett és rugalmas szerkezetű információk, mint például a JSON vagy az XML, ahol a különböző elemek különböző attribútumkészleteket tartalmazhatnak.

Az entitás a valós világ olyan konkrét vagy absztrakt tárgya, amely egyedileg azonosítható, leírható és adat formájában reprezentálható.

Felügyelt tanulás - A gépi tanulás egy olyan típusa, amelyben egy algoritmust olyan felosztott adatokon képeznek ki, ahol a kívánt eredmény minden egyes példa esetében ismert.

A taxonómia egy hierarchikus osztályozási rendszer, amelyet az elemek közös jellemzők alapján történő szisztematikus kategorizálására használnak.

A Titanic adathalmaz egy népszerű adathalmaz gépi tanulási modellek képzéséhez és teszteléséhez.

Képzés - Az a folyamat, amelynek során egy gépi tanulási algoritmus adatokat elemez, hogy mintákat azonosítson és modellt alkossan.

A transzfer tanulás egy olyan gépi tanulási technika, amelyben az egyik feladatra kiképzett modellt egy másik feladat kiindulópontjaként használják.

Transzformáció - Az adatok formátumának, szerkezetének vagy tartalmának megváltoztatása későbbi felhasználás céljából.

Adatkövetelmények - az üzleti folyamatok támogatásához szükséges információk szerkezetét, formátumát, teljességét és minőségét meghatározó formalizált kritériumok.

Az építőipar Uberizációja az építőipar hagyományos üzleti modelljeinek átalakulása az ügyfelek és a vállalkozók közötti közvetlen, közvetítők nélküli interakciót biztosító digitális platformok hatására.

Az Uniclass az Egyesült Királyságban széles körben használt második és harmadik generációs épületelem-osztályozási rendszer.

Az USD (Universal Scene Description) egy számítógépes grafikához kifejlesztett adatformátum, de egyszerű felépítése és a geometriai maguktól való függetlensége miatt a mérnöki rendszerekben is alkalmazásra került.

Az adathitelesítés az információk ellenőrzésének folyamata a megállapított kritériumok és követelmények alapján, az adatok pontosságának, teljességének és következetességének biztosítása érdekében.

Vektoradatbázis - Az adatbázisok egy speciális típusa, amely az adatokat többdimenziós vektorokként tárolja a hatékony szemantikus keresés és objektum-összehasonlítás érdekében.

A vektorreprezentáció (beágyazás) az adatok többdimenziós numerikus vektorokká történő átalakításának módszere, amely lehetővé teszi a gépi algoritmusok számára az információk hatékony feldolgozását és elemzését.

A VectorOps egy olyan módszertan, amely a többdimenziós vektoros adatok feldolgozására, tárolására és elemzésére összpontosít, különösen fontos az olyan területeken, mint a digitális ikrek és a szemantikus keresés.

Vizualizáció - Az adatok grafikus megjelenítése az információk jobb érzékelése és elemzése érdekében.

A kifejezések betűrendes kategorizálása az angol nyelvű nevük alapján történt.

HIVATKOZÁSOK ÉS ONLINE ANYAGOK LISTÁJA

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics," 2016. december 12. [Online]. Elérhető: <https://www.gartner.com/en/documents/3530919>. [A hozzáférés dátuma: 2025. március 1.].
- [2] KPMG, "Ismerős kihívások - új megközelítések. Globális építőipari felmérés 2023", 2023. január 1. [Online]. Elérhető: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Hozzáférés: 2025. március 5.].
- [3] F. R. Barnard, "Egy kép többet ér ezer szónál", 10 Mapi 1927. [Online]. Elérhető: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [A hivatkozás dátuma: 2025. március 15.].
- [4] M. Bastian, "Microsoft CEO Satya Nadella says self-claiming AGI is 'nonsense benchmark hacking'", 21 Feb. 2025. [Online]. Elérhető: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [A hozzáférés dátuma: 2025. március 15.].
- [5] W. E. Fórum, "Az építőipar jövőjét alakító fórum - átalakulóban lévő táj:", 2016. január 1. [Online]. Elérhető: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [A hozzáférés dátuma: 2025. március 2.].
- [6] C. D. Gillespie, "Agyag: a Föld összefonódása az agyag korában", 2024. [Online]. Elérhető: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papirusz a Kr. e. 3. századból, nyelve görög", 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Monitoring: a rendelkezésre álló eszközök felhasználása", 1980. [Online]. Elérhető: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [A megszólalás dátuma: 2025. március 15.].
- [9] PWC, "Adatvezérelt Amire a diákoknak szükségük van ahhoz, hogy sikeresek legyenek a gyorsan változó üzleti világban", 2015. február 15. [Online]. Elérhető: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [A hozzáférés dátuma: 2025. március 15.].
- [10] Skanska USA, "Őszi építőipari piaci trendek", 2023. november 2. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Oxford Essential Quotations (4. kiadás), Oxford University Press, 2016. [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236>. [A hozzáférés dátuma: 2025. március 1.].
- [12] "Idézet: Sondergaard az adatelemzésről," [Online]. Elérhető: <https://www.causeweb.org/cause/resources/library/r2493>. [A hozzáférés dátuma: 2025. március

- 15.].
- [13] "Hogyan lendíti fel a globális AI-érdeklődés az adatkezelési piacot", 2024. május 28. [Online]. Elérhető: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [A hozzáférés dátuma: 2025. március 15.].
- [14] И. McCue, "ERP History", 2024. [Online]. Elérhető: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "ERP ár: Mennyibe kerül egy ERP rendszer?", [Online]. Elérhető: <https://erpscout.de/en/erp-costs/>. [A hozzáférés dátuma: 2025. március 15.].
- [16] softwarepath, "Mit mond 1384 ERP-projekt az ERP kiválasztásáról (2022 ERP report)," 2022. január 18. [Online]. Elérhető: <https://softwarepath.com/guides/erp-report>. [A hozzáférés dátuma: 2025. március 15.].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects" (Adatvezérelt irányítás a digitális tőkeprojektekben), 2016. december 16. [Online]. Elérhető: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [A cím dátuma: 2025. március 1.].
- [18] McKinsey, "The data-driven enterprise of 2025", 2022. január 28. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Hivatkozás dátuma: 2024. május 22.].
- [19] Wikipedia, "Moore törvénye", [Online]. Elérhető: https://en.wikipedia.org/wiki/Moore%27s_law. [A hozzáférés dátuma: 2025. március 15.].
- [20] Accenture, "Building More Value With Capital Projects" (Több értéket teremteni a tőkeprojektekkel), 2020. január 1. [Online]. Elérhető: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accidenture-industryx-building-value-capital-projects-highres.pdf>. [A hozzáférés dátuma: 2024. március 3.].
- [21] B. Marr, "Mennyi adatot hozunk létre naponta? Az észbontó statisztikák, amelyeket mindenki el kellene olvasnia", 2018. [Online]. Elérhető: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read>.
- [22] "Mennyi adat keletkezik naponta?", 2024. [Online]. Elérhető: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?", 2024. [Online]. Elérhető: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Nyugat-Európa különböző régióiban 1454 és 1800 között fél évszázadonként előállított

- nyomtatott könyvek száma" [Internet]. [Online]. Elérhető: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [A hivatkozás dátuma: 2025. március 1.].
- [25] "Árképzési példák", 2024. [Online]. Elérhető: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Enterprises outsource data storage as complexity rises," 10 May 2024. [Online]. Elérhető: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [A hozzáférés dátuma: 2025. március 15.].
- [27] JETSOFTPRO, "A SaaS halott? Microsoft CEO's Shocking Prediction Explained," 2025. január 13. [Online]. Elérhető: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 2024. december 12. [Online]. Elérhető: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [A hozzáférés dátuma: 2025. március 15.].
- [29] GoodReads, "Tim Berners-Lee," [Online]. Elérhető: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [A hozzáférés dátuma: 2025. március 15.].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 2023. január 1. [Online]. Elérhető: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Hozzáférés: 2025. március 5.].
- [31] I. Deininger, B. Koch, R. Bauknecht és M. Langhans, "Digitális modellek használata egy termelési helyszín dekarbonizálásához: példa egy épületmodell, egy termelési modell és egy energiamodell összekapcsolására", 2024. [Online]. Elérhető: [https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Pro...duction_Site_A_Case_Stud...Connecting_the_Building_Model_Production_Model_and_Energy_Mo...del](https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Stud...).
- [32] McKinsey, "REINVENTING CONSTRUCTIONING: A ROUTE TO HIGHER PRODUCTIVITY," 2017. február 1. [Online]. Elérhető: <https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf>.
- [33] Építőipari munkacsoport a miniszterelnök-helyettesnek, "Az építőipar újragondolása", 2014. október 1. [Online]. Elérhető: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] Forbes, "Without An Opinion, You're Just Another Person With Data," 2016. március 15. [Online]. Elérhető: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [A hozzáférés dátuma: 2025. március 15.].
- [35] Wikiquote, "Charles Babbage," [Online]. Elérhető: https://en.wikiquote.org/wiki/Charles_Babbage. [A hozzáférés dátuma: 2025. március 15.].

- [36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves," 12 Mar. 2025. március 12. [Online]. Elérhető: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [A hozzáférés dátuma: 2025. március 15.].
- [37] A Kanadai Építőipari Szövetség és a kanadai KPMG, 2021, "Építőipar a digitális világban", 2021. május 1. [Online]. Elérhető: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [A hozzáférés dátuma: 2025. március 5.].
- [38] ZCS, "Az ötödik ipari forradalom megfejtése". [Online]. Elérhető: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [A hozzáférés dátuma: 2025. március 15.].
- [39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.
- [40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "A nyílt forráskódú szoftverek értéke", 2024. jan. 24. [Online]. Elérhető: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [A hozzáférés dátuma: 2025. március 15.].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook" (Szoftverfejlesztési költségbecslési kézikönyv), 2008. szeptember 1. [Online]. Elérhető: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Az építőipari termelékenység javítása" [Online]. Elérhető: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [A cím dátuma: 2025. március 15.].
- [43] A. G. a. C. Syverson, "A termelékenység furcsa és szörnyű útja az amerikai építőipari ágazatban", 2023. jan. 19. [Online]. Elérhető: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [A hozzáférés dátuma: 2025. március 1.].
- [44] McKinsey, "Delivering on construction productivity is no longer optional", 2024. augusztus 9. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [A hozzáférés dátuma: 2025. március 5.].
- [45] ING Group, "Lagging productivity in construction is driving up construction costs", 2022. december 12. [Online]. Elérhető: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [A megszólalás dátuma: 2025. március 15.].
- [46] M. Berman, "Microsoft CEO's Shocking Prediction: 'Agents Will Replace ALL Software,'" 19 December 2024. [Online]. Elérhető: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [A hozzáférés dátuma: 2025. március 15.].

- [47] Business Insider, "Anthropic CEO says that in 3 to 6 months, AI will be writing 90% of the code software developers were responsible of", 2025. március 15. [Online]. Elérhető: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [A hozzáférés dátuma: 2025. márc. 30.].
- [48] Statista, "Az adatbázis-kezelő rendszerek (DBMS) népszerűségének összehasonlítása világszerte 2024 júniusától, kategóriánként", 2024. június. [Online]. Elérhető: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [A címzés dátuma: 2025. március 15.].
- [49] DB-Engines, "DB-Engines rangsor", [Online]. Elérhető: <https://db-engines.com/en/ranking>. [Hivatkozás dátuma: 2025. március 15.].
- [50] "Stack Overflow Developer Survey 2023," 2024. [Online]. Elérhető: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Elérhető: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Strukturált és strukturálattalan adatok: mi a különbség?", 2024. [Online]. Elérhető: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "ÉPÍTÉSI PROJEKTEK ADATFORMÁTUMÁNAK PDF-VÁLTOZÁSA", 2024. ápr. 23. [Online]. Elérhető: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Building Information Modeling Whitepaper site," 2003. [Online]. Elérhető: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Boyko, "Lobbiháborúk és a BIM fejlesztése. 5. rész: A BlackRock minden technológia ura. Hogyan ellenőrzik a vállalatok a nyílt forráskódot", 2024. [Online]. Elérhető: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ushakov, "Közvetlen modellezés - kinek és miért van rá szüksége? A versenyképes technológiák áttekintése", 2011. 11. 14. [Online]. Elérhető: https://isicad.net/articles.php?article_num=14805. [A megszólalás dátuma: 2025. 02.].
- [57] C. Eastman és A. Cthers, "Eastman, Charles; And Cthers", 1974. szeptember. [Online]. Elérhető: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [A hozzáférés dátuma: 2025. március 15.].
- [58] D. Ushakov, "Közvetlen modellezés - kinek és miért van rá szüksége? A versenyképes technológiák áttekintése", 2011. november 11. [Online]. Elérhető: https://isicad.net/articles.php?article_num=14805. [A hivatkozás dátuma: 2025. március 15.].

- [59] D. Weisberg, "A CAD története", 2022. december 12. [Online]. Elérhető: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [A hivatkozás dátuma: 2025. március 15.].
- [60] ADSK, "Fehér könyv az épületinformációs modellezésről", 2002. [Online]. Elérhető: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [A hivatkozás dátuma: 2025. március 15.].
- [61] ADSK, "Fehér könyv: Épületinformációs modellezés a gyakorlatban" [Online]. Elérhető: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [A hozzáférés dátuma: 2025. március 15.].
- [62] A. Boyko, "Lobbiháborúk és a BIM fejlesztése. Második rész: nyílt BIM VS zárt BIM. Európa VS a világ többi része", 2024. [Online]. Elérhető: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
- [63] A. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs," 2024. [Online]. Elérhető: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "Whitepaper BIM", 2002. [Online]. Elérhető: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [A cím dátuma: 2025. március 15.].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Elérhető: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [A hozzáférés dátuma: 2025. március 15.].
- [66] M. Shacklett, "Strukturált és strukturálatlan adatok: 2024. [Online]. Elérhető: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Elérhető: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [Online]. Elérhető: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf> . [A hozzáférés dátuma: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Elérhető: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [A hozzáférés dátuma: 2025. március 15.].
- [70] Analyticsindiamag, "6 leginkább időigényes feladat az adattudósok számára", 2019. május 15. [Online]. Elérhető: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.

- [71] BizReport, "Report: Data scientists spend bulk of time cleaning up," 2015. július 06. [Online]. Elérhető: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [A hozzáférés dátuma: 2025. március 5.].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 2015. július 27. [Online]. Elérhető: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/. [A hozzáférés dátuma: 2025. március 15.].
- [73] B. Cyphers és K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Elérhető: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information" (Nyílt adatok: innováció és teljesítmény felszabadítása folyékony információkkal), 2013. október 1. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [A hozzáférés dátuma: 2025. március 15.].
- [75] A. Boyko, "A nyílt adatokért folytatott küzdelem az építőiparban. Az AUTOLISP, az intelliCAD, az openDWG, az ODA és az openCASCADE története", 15 05 2024. [Online]. Elérhető: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [A megszólalás dátuma: 16 02 2025].
- [76] Wikipedia, "Microsoft és a nyílt forráskód", [Online]. Elérhető: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [A hozzáférés dátuma: 2025. március 15.].
- [77] TIME, "A nyílt és zárt AI-modellek közötti szakadék csökkenhet. Here's Why That Matters", 2024. november 5. [Online]. Elérhető: <https://time.com/7171962/open-closed-ai-models-epoch/>. [A hozzáférés dátuma: 2025. március 15.].
- [78] The Verge, "More than a quarter of new code at Google is generated by AI," 2024. október 29. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [A hozzáférés dátuma: 2025. március 15.].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing," 2020. december 15. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-feldolgozás>. [Hozzáférés: 2025. március 15.].
- [80] PwC, "PwC Open Source Monitor 2019", 2019. [Online]. Elérhető: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Hozzáférés: 2025. márc. 15.].
- [81] Travers Smith, "A nyílt titok: nyílt forráskódú szoftverek", 2024. [Online]. Elérhető: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source>

- software/. [A hozzáférés dátuma: 2025. március 15.].
- [82] Deloitte, "Az adatátadási folyamat a vállalati átalakulásokban", 2021. [Online]. Elérhető: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Hivatkozás dátuma: 2025. március 15.].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery", 2024. március 20. [Online]. Elérhető: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [A hozzáférés dátuma: 2025. március 5.].
- [84] "Idézet Origin: Mindent a lehető legegyszerűbbé kell tenni, de nem egyszerűbbé", 2011. május 13. [Online]. Elérhető: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [A hozzáférés dátuma: 2025. március 15.].
- [85] "Transformer (mélytanulási architektúra)," [Online]. Elérhető: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [A hozzáférés dátuma: 2025. március 15.].
- [86] "Python csomagok letöltési statisztikái," 2024. [Online]. Elérhető: <https://www.pepy.tech/projects/pandas>.
- [87] Interview Bit, "Top 10 Python könyvtár", 2023. [Online]. Elérhető: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>. [A hozzáférés dátuma: 2025. március 30.].
- [88] "NVIDIA és HP Supercharge Data Science and Generative AI on Workstations", 2025. március 7. [Online]. Elérhető: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [A hozzáférés dátuma: 2025. március 15.].
- [89] P. Orac, "Hogyan dolgozzunk fel egy több millió soros DataFrame-et másodpercek alatt", 2024. [Online]. Elérhető: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Mi az a Kaggle?", 2024. [Online]. Elérhető: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "Az NVIDIA vezérigazgatója, Jensen Huang Keynote a COMPUTEX 2024-en", 2024. június 2. [Online]. Elérhető: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Hozzáférés 2025. március 15.].
- [92] "Tagok: Alapítók és vállalati tagok", 2024. [Online]. Elérhető: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "A változás kora: az IFC a múlté, avagy miért hajlandó az ADSK és más CAD-gyártók feladni az IFC-t az USD-ért 14 kulcsfontosságú tény," 2024. november 24. [Online]. Elérhető:

- <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Hozzáférés: 2025. február 23.].
- [94] A. Boyko, "A BIM utáni világ. Átmenet az adatokra és folyamatokra, és hogy az építőiparnak szüksége van-e szemantikára, formátumokra és interoperabilitásra", 2024. dec. 20. 20. [Online]. Elérhető: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [A megszólalás dátuma: 2025. február 23.].
- [95] N. I. o. Egészségügy, "NIH STRATÉGIAI TERV AZ ADATTUDOMÁNYOK TUDÁSÁRA", 2016. [Online]. Elérhető:
https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [A hozzáférés dátuma: 2025. február 23.].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year", 2016. szeptember 22. [Online]. Elérhető: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Az adatminőség hatásai", 2025. január 1. [Online]. Elérhető: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects" (Tervezés a karbantarthatóság érdekében: az üzemeltetési és karbantartási szempontok fontossága az építési projektek tervezési szakaszában), [Online]. Elérhető: <https://www.wbdg.org/resources/design-for-maintainability>. [A hozzáférés dátuma: 2025. március 15.].
- [99] O. o. D. C. P. a. Oversight, "Korróziómegelőzési és -ellenőrzési tervezési útmutató katonai rendszerek és berendezések számára", 2014. április. [Online]. Elérhető:
<https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%202014.pdf>. [A hozzáférés dátuma: 2025. március 15.].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights" (Adatminőség: Legjobb gyakorlatok a pontos meglátásokért), 2025. január 1. [Online]. Elérhető: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "For Want of a Nail," [Online]. Elérhető: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [A hozzáférés dátuma: 2025. március 15.].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 2013. október. [Online]. Elérhető:
https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [A hozzáférés dátuma: 2025. március 15.].
- [103] EY, "A karbonsemlegességhoz vezető út", 2023. március 10. [Online]. Elérhető:
https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [A hozzáférés dátuma: 2025. március 15.].

- [104] PWC, "ESG tudatosság", 2024. július 1. [Online]. Elérhető: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [A hozzáférés dátuma: 2025. március 15.].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Elérhető: <https://greenbuildingencyclopedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO₂_számítás a megtestesült szén-dioxidra", 2024. [Online]. Elérhető: https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon.
- [107] McKinsey, "Imagining Construction's Digital Future" (Az építőipar digitális jövőjének elképzelése), 2016. június 24. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [A hozzáférés dátuma: 2025. február 25.].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 2024. október 10. [Online]. Elérhető: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [A címzés dátuma: 2025. március 15.].
- [109] SAS, "Data lake és data warehouse - ismerd meg a különbséget", [Online]. Elérhető: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [A hivatkozás dátuma: 2025. március 15.].
- [110] ADSK, "Épületinformációs modellezés", 2002. [Online]. Elérhető: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [A cím dátuma: 2025. március 15.].
- [111] A. Boyko, "BIM történeti térkép", 2024. [Online]. Elérhető: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "A BIM definíciói szervezetek és szabványok szerint", 2023. december 27. [Online]. Elérhető: <https://encyclopedia.pub/entry/53149>. [A megszólalás dátuma: 2025. március 5.].
- [113] CAD Vendor, "OPEN BIM program", 2012. [Online]. Elérhető: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [A hivatkozás dátuma: 2025. március 30.].
- [114] Wikipedia, "Ipari alapozó osztályok", [Online]. Elérhető: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [A hozzáférés dátuma: 2025. március 15.].
- [115] Wikipedia, "IGES," [Online]. Elérhető: <https://en.wikipedia.org/wiki/IGES>. [A hivatkozás dátuma: 2025. március 30.].
- [116] A. Boyko, "A CAD (BIM) története", 2021. december 15. [Online]. Elérhető: https://miro.com/app/board/o9J_laML2cs=/. [A címadás dátuma: 2025. február 24.].

- [117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Referenciatanulmány az IFC szoftvertámagatásról: a GeoBIM benchmark 2019 - I. rész", 2021. jan. 8. [Online]. Elérhető: <https://arxiv.org/pdf/2007.10951.pdf>. [A hozzáférés dátuma: 2025. március 5.].
- [118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM", 2021. április 13. [Online]. Elérhető: <https://www.youtube.com/watch?v=VO3Y9uzF9M&feature=youtu.be>. [A megszólalás dátuma: 2025. március 5.].
- [119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 2024. december 17. [Online]. Elérhető: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [A hozzáférés dátuma: 2025. március 5.].
- [120] "Flex token költség", 2024. [Online]. Elérhető: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "Forget BIM and democratise access to data (17. Kolloquium Investor - Hochschule - Bauindustrie)," 2024. [Online]. Elérhető: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bauindustrie.pdf>.
- [122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh, and F. Plaschke, "Solving the construction industry productivity puzzle," 2015. [Online]. Elérhető: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industries-productivity-puzzle>.
- [123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle," 2018. január 1. [Online]. Elérhető: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [A hozzáférés dátuma: 2 Mar 2025].
- [124] Apple.com, "Pixar, Adobe, Apple és NVIDIA Alliance for OpenUSD, hogy a 3D tartalmakra vonatkozó nyílt szabványokat előmozdításák", 2023. augusztus 1. [Online]. Elérhető: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Hozzáférés: 2025. március 2.].
- [125] AECmag, "Az ADSK granuláris adatstratégiaja", 2024. július 25. [Online]. Elérhető: <https://aecmag.com/technology/autodesk-granular-data-strategy/>. [A hozzáférés dátuma: 2025. március 15.].
- [126] A. Boyko, "A változás kora: az IFC a múlté, avagy miért hajlandó az ADSK és más CAD-gyártók feladni az IFC-t az USD-ért, 14 kulcsfontosságú tény", 2024. 11. 24. [Online]. Elérhető: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [A megszólalás dátuma: 23 fevral 2025].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Csata az adatokért és az LLM és a ChatGPT alkalmazásáért az építőiparban", 2024. augusztus 7. [Online]. Elérhető: ENG BIM Cluster 2024 | Az LLM és a ChatGPT

- adat- és alkalmazási csatája az építőiparban. [Hozzáférés: 2025. március 15.].
- [128] "Jeffrey Zeldman bemutatja", 2008. május 6. [Online]. Elérhető: <https://zeldman.com/2008/05/06/content-precedes-design/>. [A hozzáférés dátuma: 2025. március 15.].
- [129] A. Boyko, "DWG elemzés a ChatGPT-vel | DataDrivenConstruction," 5 Mar. 2024. [Online]. Elérhető: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Hivatkozás dátuma: 2025. március 15.].
- [130] McKinsey , "The McKinsey guide to outcompeting in the age of digital and AI", 2023. [Online]. Elérhető: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [A megszólalás dátuma: 2025. március 30.].
- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs," 2016. március 31. [Online]. Elérhető: <https://www.forbes.com/sites/brentlykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [A hozzáférés dátuma: 2025. március 15.].
- [132] J. Bertin, "Grafika és grafikus információfeldolgozás", 2011. szeptember 8. [Online]. Elérhető: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_x_nm4tcC&redir_esc=y. [A hozzáférés dátuma: 2025. március 15.].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking" [Online]. Elérhető: <https://www.causeweb.org/cause/resources/library/r1266>. [A hozzáférés dátuma: 2025. március 15.].
- [134] Ministrymagazin, "Hogyan fedezte fel a tudomány a teremtést", 1986. január. [Online]. Elérhető: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [A hivatkozás dátuma: 2025. március 15.].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now" (Adatvezérelt átalakulás: gyorsítsuk fel a méretarányokat most), 2017. május 23. [Online]. Elérhető: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Hozzáférés: 2024. május 15.].
- [136] "Hogyan építünk adatarchitektúrát az innováció ösztönzésére - ma és holnap", 2020. június 3. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [A megszólalás dátuma: 2025. március 15.].
- [137] Oxford, "Woodrow Wilson 1856-1924", Oxford, "Woodrow Wilson 1856-1924". [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630>. [A beszéd dátuma: 2025. március 15.].
- [138] "Convertorok", 2024. [Online]. Elérhető: <https://datadrivenconstruction.io/index.php/convertors/>.

- [139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalize?", 2017. január 1. [Online]. Elérhető: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [A hozzáférés dátuma: 2025. február 18.].
- [140] "Épülő csővezeték", 2024. [Online]. Elérhető: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedia, "Apache NiFi", 2025. január 1. [Online]. Elérhető: https://de.wikipedia.org/wiki/Apache_NiFi. [A hozzáférés dátuma: 2025. március 5.]
- [142] n8n, "Gmail AI Auto-Responder: Választervezetek létrehozása a bejövő e-mailekre", 2024. május 1. [Online]. Elérhető: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Hozzáférés 2025. március 15.]
- [143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail," 2025. március 1. [Online]. Elérhető: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [A hozzáférés dátuma: 15 Mar 2025].
- [144] B. T. O'Neill, "Failure rates for analytics, AI, and big data projects = 85% - yikes!", 1 Jan. 2025. [Online]. Elérhető: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.
- [146] T. J. S. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman," Quantitative Economics with Julia, 1 Jan. 2025. [Online]. Elérhető: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [A hozzáférés dátuma: 2024. május 1.]
- [147] T. Landsall-Welfair, A nemzet aktuális hangulatának előrejelzése, Significance, 2012.
- [148] A. Boyko, "San Francisco. Építőipar 1980-2019", 2024. [Online]. Elérhető: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects," 2024. [Online]. Elérhető: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Felkészülés a jövőre a hálózatba kapcsolt építőiparral", CFMA, "Felkészülés a jövőre a hálózatba kapcsolt építőiparral". [Online]. Elérhető: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [A hozzáférés dátuma: 2025. március 15.]
- [151] Cisco, "Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing" (Cisco felmérés szerint az IoT-projektek közel háromnegyede kudarcot vall), 2017. május 22. [Online]. Elérhető: <https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to>

- three-fourths-of-iot-projects-are-failing.html.
- [152] "Növényi fossziliák megőrzéséhez szükséges feltételek", 2024. [Online]. Elérhető: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "BlackRock Fink a kötvényekről, a fúziókról és felvásárlásokról, az amerikai recesszióról és a választásokról: Teljes interjú, 2023. [Online]. Elérhető: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 híres mesterséges intelligencia katasztrófa", 2024. október 02. [Online]. Elérhető: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [A hozzáférés dátuma: 2025. március 15.].
- [155] G. Kaszparov, Deep Thinking, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle", 2025. január 1. [Online]. Elérhető: <https://en.wikipedia.org/wiki/Kaggle>. [A hozzáférés dátuma: 2025. március 15.].
- [157] Kaggle, "Titanic - gépi tanulás a katasztrófából", 2025. január 1. [Online]. Elérhető: <https://www.kaggle.com/competitions/titanic/overview>. [Hozzáférés: 2025. március 10.].
- [158] Ш. Johri, "A ChatGPT létrehozása: az adatoktól a párbeszédig", 2024. [Online]. Elérhető: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "Néhány hasznos dolog, amit a gépi tanulásról tudni kell", 2024. [Online]. Elérhető: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Idézhető idézet" [Online]. Elérhető: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [A hozzáférés dátuma: 2025. március 17.].
- [161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models," 19 December 2024. [Online]. Elérhető: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Hozzáférés 2025. március 25.].
- [162] "NVIDIA Announces Major Release of Cosmos World Foundation Models and Physical AI Data Tools," 18 Mar 2025. március 18. [Online]. Elérhető: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [A hozzáférés dátuma: 2025. március 25.].
- [163] NVIDIA, "NVIDIA Isaac Sim", [Az interneten]. Elérhető: <https://developer.nvidia.com/isaac/sim>. [Hozzáférés: 2025. március 25.].
- [164] M. Quarterly, "Why digital strategies fail", 2018. január 25. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [A

- hozzáférés dátuma: 2025. március 15.].
- [165] M. J. Perry, "Kedvenc Milton Friedman-idézeteim", 2006. november 17. [Online]. Elérhető: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [A hozzáférés dátuma: 2025. március 1.].
- [166] J. A. Wheeler, "Információ, fizika, kvantum: a kapcsolatok keresése", 1990.
- [169] A. Boyko, "Lobbiháborúk és a BIM fejlesztése. 5. rész: A BlackRock minden technológia ura. Hogyan ellenőrzik a vállalatok a nyílt forráskódot", 2024. [Online]. Elérhető: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen és J. Beetz, "A SPARQL query engine for binary-formatted IFC building models," *Advanced Engineering Informatics*, 2024.
- [171] "Az építőipari vállalkozások száma az Egyesült Királyságban 2021-ben, vállalkozásméret szerint", 2024. [Online]. Elérhető: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5000 IFC&RVT projekt", 2024. [Online]. Elérhető: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] M. Popova, "A bitból: John Archibald Wheeler úttörő fizikus az információról, a valóság természetéről és arról, hogy miért élünk egy részvételi univerzumban", 2008. [Online]. Elérhető: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [A hivatkozás dátuma: 2025. február].
- [174] *Lobbiháborúk az építőipari adatokért / Techno-Feudalizmus és a BIM rejtett múltjának története*. [Film]. Németország: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Dokumentumok és adatok automatikus visszakeresése a projektekből", 2023. november 16. [Online]. Elérhető: https://www.youtube.com/watch?v=ASXolti_YPs&t. [A hozzáférés dátuma: 2025. március 2.].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value's", 2022. jan. 22. [Online]. Elérhető: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [A hozzáférés dátuma: 2025. február 15.].
- [177] KPMG, "Építőipar a digitális világban", 2021. május 1. [Online]. Elérhető: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [A hozzáférés dátuma: 2024. április 5.].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 2023. márc. 17. [Online]. Elérhető: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [A hozzáférés dátuma: 2025. február 15.].

- [179] O. Business, "Satya Nadella Reveals 'How AI Agents Will Disrupt SaaS Models'", 2025. január 10. [Online]. Elérhető: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [A hozzáférés dátuma: 2025. március 15.].
- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23 Mar. 2016. március. [Online]. Elérhető: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [A hozzáférés dátuma: 2025. március 15.].
- [181] UK Foreign, Commonwealth and Development Office, "Digital development strategy 2024 to 2030", 2024. március 18. [Online]. Elérhető: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [A hozzáférés dátuma: 2025. március 15.].
- [182] "Vízió és stratégia az épülettervezési iparban", 2003. november 7. [Online]. Elérhető: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [A beszéd dátuma: 2025. március 5.].
- [183] M. Bocharov, "Információs modellezés", 2025. március. [Online]. Elérhető: <https://www.litres.ru/book/mihail-evgenievich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [A megszólalás dátuma: 2025. március 15.].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale," 2000. [Online]. Elérhető: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [A hivatkozás dátuma: 2025. március 25.].
- [185] CAD Vendor, "Open BIM Programme is a marketing campaign," 2012. március 12. [Online]. Elérhető: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [A megszólalás dátuma: 2025. március 30.].

TÁRGYI INDEX

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477, 395, 457, 459, 461, 477.

AIA, 289

AIM, 3, 289

AMS, 14, 84, 151, 153

Apache Airflow, 311, 361, 362, 363, 364, 364, 366, 367, 369, 371, 399, 494

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494, 482, 494

Apache ORC, 62, 330, 378

Apache Parketta, 62, 67, 380, 381, 494, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488.

B

BDS, 258, 259

Big Data, 9, 67, 245, 383

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328,

337, 351, 356, 358, 361, 389, 413, 416, 448, 462, 466, 475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

Hatóroló doboz, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284, 276, 283, 284

C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421, 343, 349, 421

Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443.

CO₂, 72

CO₂, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

Európa Tanács, 56, 168, 169, 170, 477

Másodpilóta, 114, 116, 459

CPIXML, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473, 166, 175, 233, 326, 473

CQMS, 14, 84, 177, 178, 462, 177, 178, 462

CRM, 109, 369, 459.

CRUD, 51, 459

CSG, 263

CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 356, 373, 377,

- 378, 379, 380, 401, 407, 414, 474, 480
- D**
- DAE, 276, 277, 278, 280, 281, 284, 285, 414
- DAG, 362, 363, 365, 366
- Dash, 320, 336, 337.
- Adatkezelés, 373, 395, 396, 398, 400, 401
- Adat-tó, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
- Data Lakehouse, 67, 373, 386, 386, 387, 388
- Adatminimalizmus, 373, 395, 396, 400, 401
- Data Swamp, 373, 395, 397, 401
- Adattárház, 382, 383, 400
- Adat mint szolgáltatás, 487
- adatvezérelt, 50, 170, 460, 461, 484, 486, 460, 461, 484, 486
- DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475.
- DataOps, 170, 373, 398, 399, 400, 401, 475
- DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494, 429, 443, 494.
- DGN, 8, 140, 186, 227, 357, 186, 227, 357
- DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
- DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390, 387, 388, 382, 383, 384, 386, 387, 388, 389, 390
- DXF, 8, 73, 277
- E**
- ECM, 58, 175
- ECS, 142
- EIR, 289
- eLOD, 289
- ELT, 384, 385
- EPM, 14, 166, 198
- ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
- ESG, 196, 235, 236, 238
- ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
- Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 123,
- 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497.
- Kivonat, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475
- F**
- Toll, 62, 123, 330, 378
- FPDF, 339, 340, 341, 342, 343, 340, 341, 343
- G**
- GDPR, 109
- GIS, 58
- GLTF, 143, 278
- Google Sheets, 368, 370
- Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443.
- H**
- HDF5, 62, 67, 123, 329, 330, 378, 379, 380, 123, 329, 330, 378, 379, 380
- HiPPO, 29, 37, 95, 424, 477, 484, 490
- HTML, 123, 340, 365, 370
- I**
- AZONOSÍTÓK, 289, 290, 291.
- IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 417, 497.
- IGES, 262, 263, 276.
- iLOD, 289
- IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 417, 418, 419, 455, 460, 465, 482, 484.
- ISO 19650, 388
- J**
- JavaScript, 320, 369, 378
- JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
- Jupyter notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425, 425
- K**
- Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
- k-NN, 392, 393, 442, 445, 446, 447, 448, 442, 445, 446,

447, 448

KPI, 245, 311, 317, 318, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.

LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329, 343, 349, 381, 421, 429, 443, 494

LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494

Terhelés, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475.

LOD, 287, 289

LOI, 287

LOMD, 287

Alacsony kód, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494, 409, 411, 494

MCAD, 77, 78, 257, 284

MEP, 14, 175, 492

HÁLÓ, 234, 283, 284, 285, 296, 462, 284, 285, 296, 462

Microsoft SQL, 65

Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494.

MRP, 11, 12

MS Project, 70

MySQL, 63, 64, 65, 291, 329, 330, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372

NLP, 69, 394

No-kód, 368, 369

NURBS, 142, 282, 283, 284, 285

NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285

OCCT, 273

OCR, 69, 128, 132, 134, 128, 132, 134

OmniClass, 154, 155, 156.

nyitott BIM, 142, 216, 219, 256, 261, 278, 291

Nyílt forráskód, 45, 55, 97, 98, 107, 108, 273, 275, 494, 97, 98, 107, 108, 273, 275, 494

BAGOLY, 267, 268, 269.

P

Pandák, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494.

Parketta, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480

PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497.

PDM, 16

PHP, 63, 339

Csővezeték, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476

PLM, 16, 246

PLN, 8, 272, 296

Plotly, 320, 336, 337, 372

PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487.

PMS, 84, 151, 233

PostgreSQL, 63, 64, 65, 329, 395

Power BI, 320, 372, 482

magántőke, 464, 487

Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494.

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479

QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 349, 381, 429, 443

R

RAG, 111, 116

RDBMS, 63, 64, 65, 82, 89.

RDF, 267, 268, 269
 RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474, 337,
 337, 474
 RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482, 465, 482
 ROI, 311, 317, 319, 321, 351, 370, 389, 479, 479
 FORDULATSZÁM, 14, 84, 151, 331, 333, 334, 331, 333, 334
 RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329,
 357, 376, 414, 415, 417, 497, 376, 414, 415, 415, 417,
 497

S

SaaS, 24, 50, 51, 52, 458
 HATÓKÖR, 273, 277
 SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414, 264,
 273, 281, 286, 296, 329, 414
 Seaborn, 123, 320, 336, 337, 408, 412
 SPARQL, 269
 SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268,
 269, 279, 276, 277, 300, 329, 376, 392, 494
 SQLite, 63, 64, 65, 166, 167, 296, 329, 363, 166, 167, 296,
 329, 363
 LÉPÉS, 261, 262, 263, 263, 266, 272, 276, 277, 292
 Streamlit, 336, 337
 SVF, 142, 276, 285

T

Átalakítás, 128, 193, 311, 323, 325, 326, 330, 331, 333,
 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367,
 383, 384, 450, 475.

U

Uniclass, 154, 155, 156.
 USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285,
 286, 292, 296

V

VectorOps, 373, 398, 400, 401
 VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296,
 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
 XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291,
 299, 329, 378, 414, 474, 329, 128, 143, 269, 272, 277,
 280, 291, 299, 329, 378, 414, 474.